DOCUMENT RESUME

ED 429 284 CS 013 543

TITLE	Michigan High School Proficiency Test in Communication Arts: Reading. Tryout and Pilot Technical Report.
INSTITUTION	Michigan State Dept. of Education, Lansing. Michigan Educational Assessment Program.
PUB DATE	1998-03-00
NOTE	101p.; For related tests, see ED 428 944-945.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	High Schools; Pilot Projects; *Reading Achievement; *Reading Tests; Student Writing Models; Tables (Data); *Test Construction; *Test Format
IDENTIFIERS	Michigan Educational Assessment Program

ABSTRACT

As part of the test development process, this technical report is intended to present technical information from the tryout and the pilot stages of the Michigan High School Proficiency Test (HSPT) in Communication Arts: Reading. There are four major parts to this report. Part 1, Evolution of the HSPT in Communication Arts: Reading, introduces the purpose, the legislation, and the committees involved in test development. Development of the reading assessment framework and the structure of the framework are briefly described in this part. Part 2 provides an overview of the exercise development of the test. Part 3 summarizes the process used in sampling, the tryout design, the rating process for extended-response questions, reader reliability, test statistics and analyses, and other technical issues for the HSPT in Communication Arts: Reading tryout and pilot administrations. Summary results from student and teacher surveys conducted during the tryout stage are included in Part 4. Appendixes contain lists of committee members; checklists; scoring guides; expert panel recommendations; tryout and pilot statistics; survey instruments; and survey results. Contains 18 references. (RS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 429 284

Michigan High School Proficiency Test in Communication Arts: Reading Tryout and Pilot Technical Report

CS 013 543

Michigan Educational Assessment Program
Michigan Department of Education
March 1998

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
C. J. Skidmore
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

Table of Contents

	Page
List of Tables	iii
List of Figures	iv
Introduction	1
1. Evolution of the HSPT in Communication Arts: Reading	1
The Purpose of the Michigan High School Proficiency Test	1
The Expert Panel	1
Legislation Change	2
Developing the Assessment Framework to Guide the Development of the HSPT in Communication Arts: Reading	5
Structure and Implementation of the Assessment Framework	5
Committees Involved in the Development of the HSPT	6
The Technical Advisory Committee (TAC)	6
The Exercise Development Team (EDT)	6
The Content Advisory Committee (CAC)	6
The Bias Review Committee (BRC)	7
2. Exercise Development for the HSPT in Communication Arts: Reading	9
General Item Specifications	9
Item Response Formats	9
Test Directions	10
3. HSPT in Communication Arts: Reading Tryout and Pilot	11
Sample Design and Characteristics	11
Tryout Test Design	12
Rating Process for Extended-Response Items	14
Interrater Reliability	14
Tryout Statistics and Analyses	15
Item Difficulty	15
Test Reliability	16
Content Validity	16
Calibration Models	16
3PL/2PPC Model	17
Rasch Model	18
Calibration Analyses	18
Fit Statistics and Analyses	18
Item Discriminations	19
Equating	20
Scaling Model Selection	20
Racial and Gender Bias Analyses	21
Mantel Statistic for Ordered Response Categories	21
Standardized Mean Difference	22
Distributions of Standardized Mean Differences	23
Overall DIF Rating	23
Pilot Test	24
Pilot Sampling	25

Pilot Administration 25
General Results 25
Interrater Agreement 26
Group Descriptive Analyses 27
Gender/Ethnicity DIF Statistics 27
Summary 27

4. Student Survey and Teacher Survey 28

Communication Arts: Reading Student Survey Results 28
Conclusions from the Student Survey 29
Communication Arts: Reading Teacher Survey 29
Summary of the Teacher Survey Results 30
Overall Summary and Follow-Up 30

References 32

Appendices 33

A: Committees Assisting in the Development of the HSPT 33
Expert Panel Recommendations 38
Michigan School Stratum Classifications 42
Criteria for Writing and Editing Multiple-Choice Items 43
Checklist for Item Development 44
Checklist for Scoring Rubrics/Scoring Guide 46
Scoring Guide for Communication Arts: Reading 47
B: Tryout Statistics for the HSPT in Communication Arts: Reading 49
Teacher Comment Sheet 79
C: Pilot Statistics for the HSPT in Communication Arts: Reading 81
D: Student Survey 84
Student Survey Results 86
Teacher Survey Results 89

List of Tables

Table		Page
1	HSPT Development Timeline	4
2	Number of Sampled Schools in the Tryout by Stratum	11
3	Distribution of Students by Gender in the Tryout by Form	12
4	Distribution of Students by Ethnicity in the Tryout by Form	12
5	Configuration of the Tryout	13
6	Tryout Form Composition for Communication Arts: Reading	13
7	Interrater Agreement and Consistency Ranges for the Tryout	15
8	Raw Score Statistics by Form for the Tryout	49
9	Item Means and Standard Deviations by Form	50
10	Item P-Values of Quintiles by Form	60
11	Summary of Fit Results - 1PL/PPC	70
12	Summary of Fit Results - 3PL/2PPC	71
13	Items Flagged for Deletion Under the Fit Criteria - 1PL/PPC and 3PL/2PPC	72
14	Mean and Standard Deviations of Item Discrimination - 3PL/2PPC	73
15	Ranges of Racial and Gender SMDs in the Tryout	23
16	Frequency Distribution of Items by Racial SMDs	23
17	Frequency Distribution of Items by Gender SMDs	23
18	Overall DIF Rating Classification as a Function of Gender and Race	24
19	Frequency Distribution of Items by Overall DIF Rating	24
20	Mantel-Haenszel Statistics by Form for the Tryout	74
21	Number of Students Participating in the Pilot by Form	25
22	Descriptive Statistics by Form for the Pilot	81
23	Interrater Reliability Ranges	26
24	Mean Interrater Agreement Between the First Two Readers	26
25	Frequency of Interrater Agreement for Extended-Response Items by Form	82
26	Group Descriptive Statistics	27
27	Standardized Mean Differences for Gender and Ethnic Groups	83
28	Student Survey Results Summary	29
29	Teacher Survey Results Summary	30
30	Student Survey	84
31	Student Survey Response Means in Reading	86
32	Student Survey Mean Scores by Part	88
33	Teacher Survey - Reading: Statements with ≥ 20% Responding NT	89

List of Figures

Figure	Page
1 - HSPT Development Process	3
2 - Communication Arts: Reading Framework for Curriculum and Assessment	8
3 - Configuration of Form Triplets and Quadruplets for Equating	20

Introduction

As part of the test development process, this technical report is intended to present technical information from the tryout and the pilot stages of the Michigan High School Proficiency Test (HSPT) in Communication Arts: Reading. There are four major parts to this report. Part 1, Evolution of the HSPT in Communication Arts: Reading, introduces the purpose, the legislation, and the committees involved in test development. Development of the reading assessment framework and the structure of the framework are briefly described in this part. Part 2 provides an overview of the exercise development of the test. Part 3 summarizes the process used in sampling, the tryout design, the rating process for extended-response questions, reader reliability, test statistics and analyses, and other technical issues for the HSPT in Communication Arts: Reading tryout and pilot administrations. Summary results from student and teacher surveys conducted during the tryout stage are included in Part 4. The relevant data tables are furnished in the appendices. Operational technical reports will follow a similar format.

Part 1. Evolution of the HSPT in Communication Arts: Reading

The Purpose of the Michigan High School Proficiency Test

As required by law, The Michigan High School Proficiency Test (HSPT) was developed to provide students with an opportunity to earn state endorsement of the local diploma. Public Act 118 (P.A. 118) of 1991, Section 104(a)(subsection 7) of the School Aid Act states:

> Not later than July 31, 1993, the department shall develop and the state shall approve assessment instruments to determine pupil proficiency in communication arts, mathematics, science and other subject areas specified by the state board. The assessment instruments shall be based on the state board model core curriculum outcomes. Beginning with the graduating class of 1997, a pupil shall not receive a high school diploma unless the pupil achieves passing scores on the assessment instruments developed under this section.

The legislation initiating the development of the HSPT was introduced to respond to educators' and employers' concern that Michigan students were leaving high school without the knowledge and skills necessary to lead productive lives. Additionally, the high school diploma was awarded on the basis of local requirements. There was no consistency from school to school, nor were there, with the exception of one semester's instruction in civics, state requirements for receiving a high school diploma. The HSPT provides a consistent measure of what students should know and be able to do at the end of the tenth grade in Michigan schools.

The Expert Panel

The Expert Panel on the Michigan High School Graduation Test was convened to advise the Michigan State Board of Education on important issues surrounding the high school proficiency examination enacted by P.A. 118 of 1991. The panel consisted of national experts with first-hand knowledge and experience in large-scale testing programs (see Appendix A for list of Expert Panel members).

The Expert Panel met over three days in February and March of 1992 to examine the educational, technical, legal, fiscal and logistical issues relating to competency testing and the steps to be taken in the implementation of P.A. 118. Its report "Issues and Recommendations Regarding Implementation of the Michigan High School Graduation Tests" was issued in April of 1992. The

report included 51 recommendations and rationale for each of the recommendations (see Appendix A).

Legislation Change

Between the issuance of the Expert Panel Report and the development of the assessment frameworks for each of the content areas tested by the HSPT, new legislation was passed which dramatically changed the intent of the test. Whereas P.A. 118 had stated that the awarding and denying of high school diplomas would be determined by HSPT scores, Public Act 335 of 1993 softened the intent of the test. P.A. 335, Section 1279 states that the HSPT would be used to award state endorsements of the local high school diploma:

> Beginning with pupils scheduled to graduate in 1997, if a pupil achieves the academic outcomes required by the state board, as measured by an assessment instrument developed under subsection (8), for a state-endorsed high school diploma in 1 or more of the subject areas of communications skills, mathematics, science, and, beginning with pupils scheduled to graduate in 1999, social studies, the pupil's school district shall award a state endorsement on the pupil's diploma in each of the subject areas in which the pupil demonstrated the required proficiency. A school district shall not award a state endorsement to a pupil unless the pupil meets the applicable requirements for the endorsement, as described in this subsection. A school district may award a high school diploma to a pupil who successfully completes local district requirements established in accordance with state law for high school graduation, regardless of whether the pupil is eligible for any state endorsement... The assessment instruments shall be based on the state board model core academic curriculum outcomes...

The change in the law also changed the context in which the Expert Panel Recommendations were considered in the development of the HSPT. In addition to the Expert Panel Report, several policy decisions and subsequent policy actions shaped the development of the HSPT from the onset.

- The HSPT would align with the *Michigan Model Core Curriculum Outcomes* (State Board of Education, 1991), broad outcomes to be achieved by all students as a result of their school experiences. Fundamental to the *Model Core Curriculum Outcomes* is the belief that the ultimate purpose of education is to permit each individual student to reach his or her optimum potential, to lead a productive and satisfying life (*The Common Goals of Michigan Education*, 1980).
- The HSPT would establish high expectations for all students.
- The HSPT would focus on the application of knowledge, problem solving and critical thinking.
- The HSPT would assess what students should know and be able to do by the end of tenth grade.
- Recognizing that what gets tested, gets taught, the HSPT would, to the extent possible in large-scale assessment, model good instructional practice.

Students achieving proficient scores on the Michigan High School Proficiency Test in mathematics, science, writing and reading earn the state endorsement of the local diploma in mathematics, science and communication arts.

Table 1 and Figure 1 show the timeline and the process used by the Michigan Department of Education, Michigan Educational Assessment Program (MEAP) for the development of the HSPT.

Figure 1. HSPT Development Process

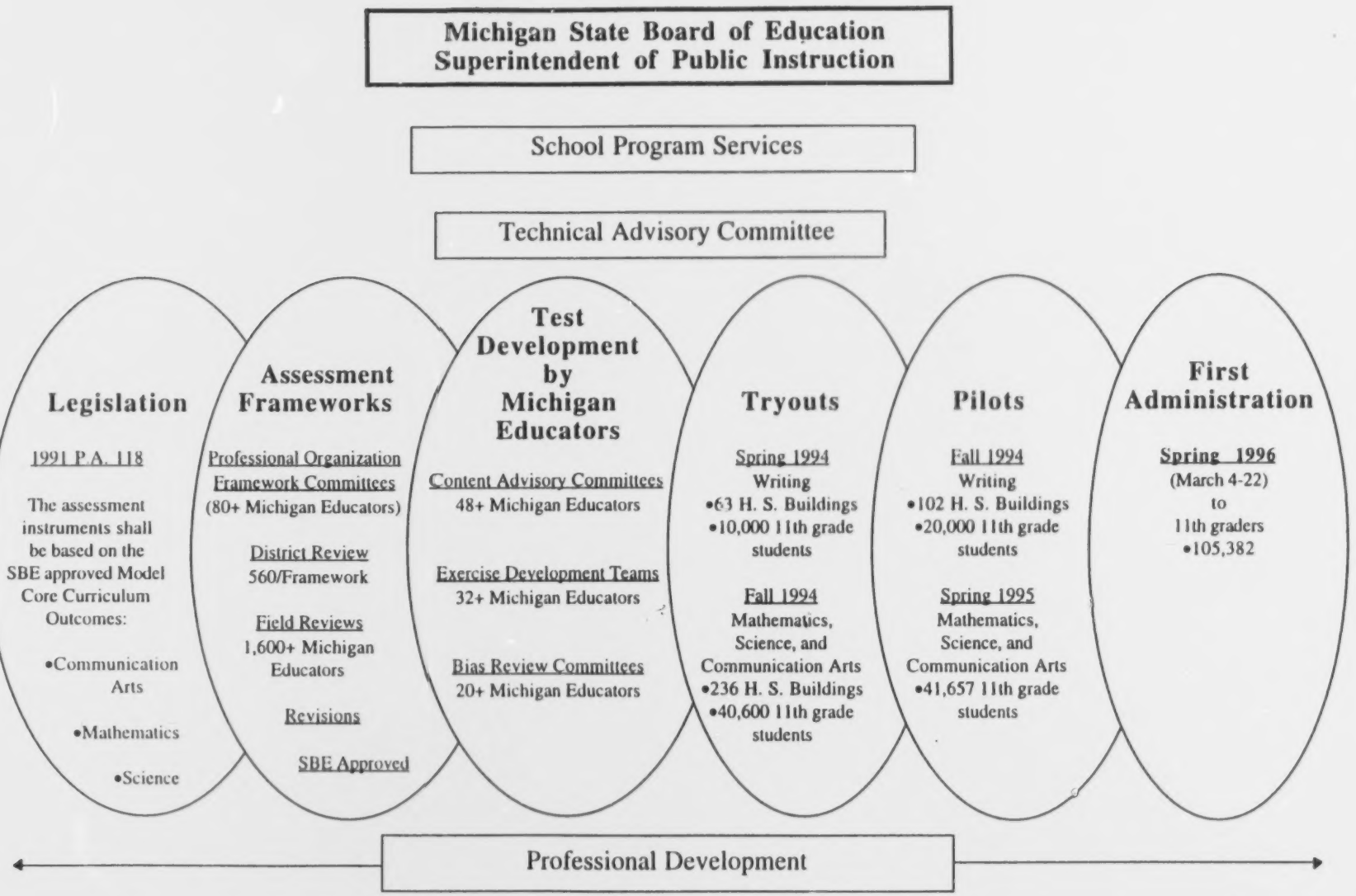


Table 1. HSPT Development Timeline

High School Proficiency Test ***Timeline 1992-1997*** Mathematics, Science, Reading, Writing	
1992-1993	**Define Test Frameworks**
November 2, 1992	Met with MRA, MSTA, MCTM and MCTE to discuss Frameworks development
January 8, 1993	Proposals to Michigan Department of Education
February, 1993	Input: Preliminary Field Review by Professional Organizations
March 31, 1993	Frameworks due to Michigan Department of Education
April 21, 1993	Michigan State Board of Education receives Frameworks
April 21 - May 31, 1993	Field Review and Comments
Summer, 1993	State Board of Education Approves Frameworks
1993, 1994, 1995	**Test Development**
Summer 1993 November 1993 January 1994	Issued RFPs Item/Exercise Development-Writing Test Item/Exercise Development- Mathematics, Science, Reading
April 1994	Tryouts-Writing Scoring, Analysis and Revision
November 1994 November 1994 April 1995	Pilots-Writing Scoring and Analysis Tryouts-Mathematics, Science, Reading Scoring, Analysis and Revision Pilots-Mathematics, Science, Reading Scoring, Analysis
1996-1997	**Test Administration Timeline**
Spring 1996	Test Administration
Fall 1996	Retest
Spring 1997	Test/Retest Award Endorsements Based Upon Results

Developing the Assessment Framework to Guide the Development of the HSPT in Communication Arts: Reading

The *Assessment Framework for the HSPT in Communication Arts: Reading* (1994) was developed by the Michigan Reading Association (MRA) to guide the development of the Michigan High School Proficiency Test in Communication Arts: Reading. The Framework reflects the reading outcomes contained in the Michigan State Board of Education *Model Core Curriculum Outcomes* (1991) and the *Michigan Essential Goals and Objectives for Reading* (1986).

The Assessment Framework was constructed from a thorough review of current research and instructional practices, discussion with state and national leaders, and field reviews. In addition, particular attention was paid to existing framework documents from other state assessment programs, the National Assessment of Educational Progress (NAEP), Council of Chief State School Officers (CCSSO) projects, and the New Standards Project to identify relevant research and resources for review. The Frameworks Project Team consisted of 13 members, including university professors, local district administrators, classroom teachers and MDE personnel. Framework development committee members were listed in the appendix of the Framework.

On April 21, 1993, the Michigan State Board of Education received the *Assessment Framework for the Michigan High School Proficiency Test in Communication Arts: Reading* and authorized the Superintendent of Public Instruction to disseminate the Framework to every school district in the state for a second round of field reviews and comments.

Structure and Implementation of the Assessment Framework

The Framework describes what students are expected to learn by the end of tenth grade, the process students use to learn, and how what is learned is assessed (see Figure 2, p. 8). Three separate assessment components were outlined in the Framework:

- standard tasks - two different products or performances produced by students at the end of an extended classroom unit of study;
- portfolios to document change over time and which might include the standard tasks; and,
- on-demand tasks.

Because of the relatively short test development timeline, it was proposed in the Framework that only the on-demand task be implemented in the tryout, pilot and 1996 operational forms of the test. During the test development process, other changes were made to the reading test as it was originally proposed in the Framework. First, the reading test was shortened from the framework version that would have required a minimum of two and one half hours administration time to a design that enabled school districts to administer the (untimed) test in two fifty-minute class periods. This was done to lessen the overall testing time of the HSPT, taking into consideration that Michigan law mandated writing, mathematics, science and, eventually, social studies be measured, in addition to reading. Secondly, some terms found in the *Model Core Curriculum Outcomes* document were substituted in test materials for less familiar framework terminology.[1] However, the characteristics of the tryout, pilot and

[1] The term "Constructing Meaning" was substituted for the framework term "Acquiring and Interpreting Knowledge" and "Knowledge About Reading" was substituted for "Metacognition." On the test, students demonstrate "Extending and Refining Knowledge" (framework) through the "Across the Reading Selection" items. Students demonstrate "Authentic Application of Knowledge" (framework) through their extended response on the "Response to the Reading Selection" test question.

operational forms of the HSPT in Communication Arts: Reading remain as described in the Framework document: (a) a common thematic focus, (b) a common focus question, (c) a variety of genre, (d) a range of difficulty levels for materials, (e) different points of view, (f) an issue or problem that persists across time, (g) diverse perspectives, (h) different lengths of materials, (i) an authentic project or performance, (j) a common scoring guide, (k) administrative standards, and (l) guidelines for time requirements.

Committees Involved in the Development of the High School Proficiency Test (HSPT)

The Technical Advisory Committee (TAC)

After the Expert Panel submitted its recommendations for implementing the HSPT, a subset of six core panel members was selected to form the Technical Advisory Committee (TAC) to serve in an advisory capacity during test development and implementation. Additional membership has been determined on an ad hoc basis as needed for particular expertise. The TAC has met with Michigan Educational Assessment Program (MEAP) staff four times or more a year to provide continuous advice on technical, policy and legal issues related to the MEAP tests.

Prior to the first meeting, each TAC member received executive summaries of the assessment frameworks in mathematics, science, reading, writing and portions of the proposal submitted by CTB/McGraw-Hill, the vendor chosen to coordinate item development for mathematics, science and reading. The TAC played an active role throughout test development and standard setting: shaping and reviewing plans, advising staff on the appropriate analyses to require of contractors and reviewing analyses provided. The TAC has been intimately involved in the program at every step and continues to be involved.

The Exercise Development Team (EDT)

The Exercise Development Team for Communication Arts: Reading was made up of ten Michigan teachers who were nominated by MDE Curriculum and MEAP staff. Members of the EDT signed a contract before item writing began. The committee members were responsible for writing all of the HSPT in Communication Arts: Reading items. All members received item writing training from CTB/McGraw-Hill. More information about exercise development for the HSPT is contained in a later section of this report.

The Content Advisory Committee (CAC)

The Content Advisory Committee for Reading was responsible for the integrity of the HSPT in Communication Arts: Reading. The CAC reviewed each test item to ensure that it was appropriately related to the *Model Core Curriculum Outcomes* (1991) and the *Michigan Essential Goals and Objectives in Reading* (1986), as set out in the legislation. Both of these documents were approved by the State Board of Education and disseminated to school districts well in advance of the first administration of the HSPT in the spring of 1996. Items were evaluated for consistency with the criteria set out in the Assessment Framework and appropriateness for measuring proficiency in reading for all students by the end of tenth grade. The CAC reviewed every test form to check for a reasonable distribution of item difficulty and for an adequate sample of the content area. Items were rejected or revised based upon decisions made by the Content Advisory Committee. However, not all forms were reviewed equally and thoroughly because of time constraints.

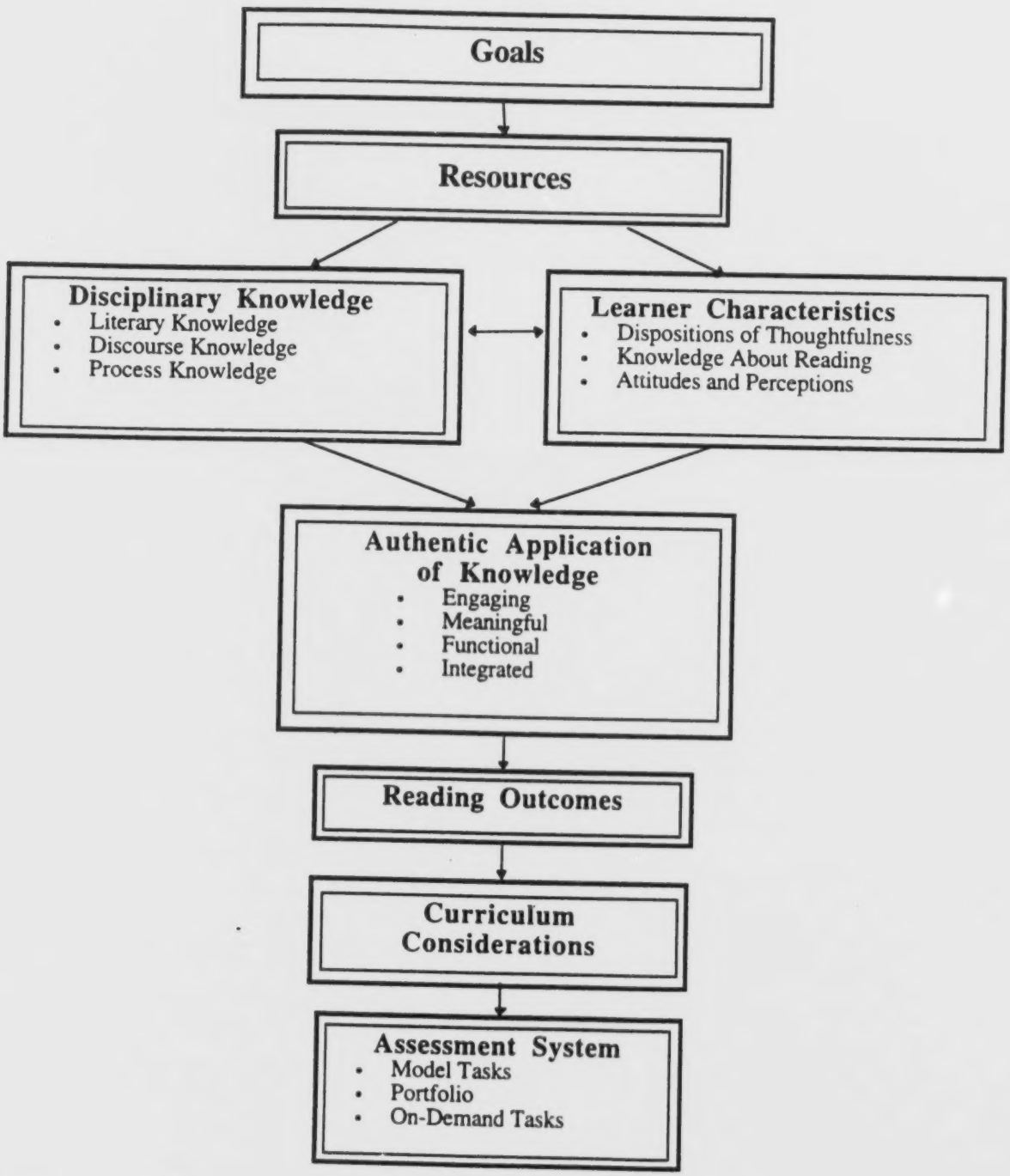
The CAC for Communication Arts: Reading was originally made up of nine members including high school and middle school classroom teachers, district and school reading department chairpersons, college reading instructors and the reading consultant from the Curriculum Development Unit of MDE.

The Bias Review Committee (BRC)

The first Bias Review Committee was comprised of eleven members from the Michigan Department of Education and several Michigan school districts. School district personnel ranged from administrators to content area consultants to English as a Second Language (ESL) coordinators and classroom teachers. BRC members reviewed every HSPT test item for possible bias to gender, racial or ethnic groups; religious groups; socioeconomic groups; persons with disabilities; older persons; and for regional concerns. In instances where the BRC observed bias, the BRC was responsible for providing suggestions that made the test material as bias-free as possible, but did not distort or interfere with test content. The BRC continues to meet with MEAP staff on a periodic basis to review new forms or as needed.

Lists of members of the above committees are in Appendix A.

Figure 2: Communication Arts: Reading Framework
For Curriculum and Assessment



Part 2. Exercise Development for the HSPT in Communication Arts: Reading

A major portion of the work in the Michigan Educational Assessment Program has been done contractually. Through the Department of Budget and Office of Purchasing, the Department of Education issues a Request for Proposals (RFP) describing the Department's testing requirements. The successful bidder must meet both quality and cost criteria as part of the evaluation process.

In order to meet the tight timeline required by legislation for development of the HSPT, CTB/ McGraw-Hill was hired to coordinate the exercise development process for the HSPT in mathematics, reading and science. CTB has years of experience in test development for national achievement tests, as well as for state assessment programs. For the HSPT, with direction from MDE Curriculum and MEAP staff, CTB provided training for the Exercise Development Team (EDT) and facilitated the EDT meetings. In addition, CTB developed the initial reading item bank and test forms and ran item analyses on the tryouts and pilot tests. The CTB contract ran through the initial pilot process.

In early 1994, notebooks were sent to all committee members of the EDT to use as a resource during the development process. The notebooks, called "The Michigan Exercise Development Guideline for Reading," contained an overall schedule for exercise development and an outline of the scope of work and specific tasks for each writer. The guidelines included general item specifications and criteria for writing and editing multiple-choice and extended-response items and for writing rubrics for the extended-response items. The EDT completed item development by June of 1994. General assessment specifications used by the reading EDT follow. Detailed specifications are contained in the Exercise Development Guidelines for Reading provided by CTB.

General Specifications

Item Response Formats

Multiple-choice items have the following requirements, in addition to the Criteria for Writing and Editing Multiple-Choice Items found in Appendix A:

- Whenever possible, the stem should be stated as a complete question, except when using an incomplete statement is clearer or less awkward.
- There must be four response options: a correct response and three incorrect responses (called foils or distractors).
- Answer choices are ordered short to long or, if single words, in alphabetical order.
- Distractors must be written with as much care and precision as the correct option, so that alternatives are attractive to a reader who does not possess the skill being assessed and therefore guesses.
- The multiple-choice items in a form should be distributed approximately equally across the reading selections.
- Items should have relevance to the focus question.

Authentic application, extended-response exercises have the following requirements:

- Each exercise will be stated as a writing prompt requiring a response of at least 1-2 pages in length.

- Each exercise will be contextualized, that is, embedded in a context, or scenario, that relates to a real-life student experience that actively engages students. The context will provide an authentic purpose for writing and will identify an audience.
- Each exercise will require students to utilize information from all three reading selections.
- Each exercise will require students to go beyond these reading selections and draw upon their own knowledge and experience.
- Each exercise will directly address the focus question.
- Each response should be scorable with a 4-point scoring rubric.

Scoring rubrics have the following requirements, in addition to those contained in the Checklist for Scoring Rubrics/Scoring Guide found in Appendix A:

- In the scoring rubric for an extended-response item, all anticipated correct responses must be concisely stated answers that will satisfy a qualified judge as being an adequate short answer to the question. The response(s) must not answer more than the question asks.
- Differences between the performance criteria must be clear and unambiguous.

Test Directions

In order to prevent readers from becoming confused when faced with multiple item formats, clear directions must be given at the beginning of each test booklet. The directions should inform the readers that there will be different item formats: one question requiring an extended written response, and 35 multiple-choice items, each with a single correct response.

Part 3. HSPT in Communication Arts: Reading Tryout and Pilot

After the Exercise Development Teams completed items for each content area to be tested on the HSPT, the Content Advisory Committees and the Bias Review Committee reviewed all items. Tryouts were scheduled for the items that survived this initial committee review. Statistical data from tryouts and pilots are part of the information used to determine which items merit further consideration for use on "live" or operational tests. In addition, participating teachers were asked to return comment sheets describing problems with the directions and/or items and noting administration details, such as the amount of time it took the majority of students to complete the test. Comments from teachers are particularly helpful in making decisions about items and test forms (see Appendix B for a sample).

Sample Design and Characteristics

Data for the HSPT in Communication Arts: Reading tryout and pilot were collected using the same procedures. To ensure representativeness, cluster sampling combined with stratification was used to sample from Michigan public schools. Michigan schools are classified into seven strata by resident population size of the community where the school is located (see Appendix A for stratum classifications). Schools participating in the tryout were randomly sampled from each stratum roughly proportional to the population proportions. The number of sampled schools in the reading tryout by stratum is listed in Table 2 below.

Table 2. Number of Sampled Schools in the Tryout by Stratum

Stratum	# of Schools Sampled	Total # of Schools in the Stratum	% of Stratum
1	4	49	8.2%
2	6	64	9.4%
3	11	106	10.38%
4	6	62	9.7%
5	0	7	0%
6	27	232	11.64%
7	19	218	8.7%
undefined[2]	7	NA	NA
Total	80	738	--

The sampled schools were considered representative of Michigan student population in gender, ethnicity, and school size. Distributions by gender and ethnic groups for the reading tryout by test form are shown in Tables 3 and 4.

Schools participating in the tryout were not sampled again for the pilot. Schools that were sampled for the tryout or pilot but did not participate were replaced by schools with similar characteristics to keep the representativeness of the sample. Also, schools participating in the reading tryout or pilot were not selected in the mathematics or science tryouts and pilots.

[2] These schools were either alternative or adult education high schools.

Table 3. Distribution of Students by Gender in the Tryout by Form

Form	Total # of Students Tested	# of Males	# of Females
1	1043	532	511
2	1035	525	510
3	950	439	511
4	983	485	498
5	884	457	427
6	1039	499	540
7	926	455	471
8	1040	498	542
9	1020	470	550
10	971	451	520
Total	9891	4811	5080

Table 4. Distribution of Students by Ethnicity in the Tryout by Form

Form	# of Students Tested	Am. Indian N (%)	Asian N (%)	Black N (%)	Hispanic N (%)	White N (%)	Multi-racial N (%)	Other N (%)
1	1043	12 (1.2)	14 (1.3)	70 (6.7)	37 (3.5)	842 (80.7)	37 (3.5)	31 (3.0)
2	1035	25 (2.4)	12 (1.2)	81 (7.8)	22 (2.1)	820 (79.2)	33 (3.2)	42 (4.1)
3	950	24 (2.5)	18 (1.9)	117 (12.3)	32 (3.4)	688 (72.4)	40 (4.2)	31 (3.3)
4	983	12 (1.2)	18 (1.8)	141 (14.3)	19 (1.9)	712 (72.4)	31 (3.2)	50 (5.1)
5	884	16 (1.8)	12 (1.4)	87 (9.8)	23 (2.6)	667 (75.5)	29 (3.3)	50 (5.7)
6	1039	10 (1.0)	13 (1.3)	83 (8.0)	11 (1.1)	849 (81.7)	27 (2.6)	46 (4.4)
7	926	18 (1.9)	16 (1.7)	82 (8.9)	16 (1.7)	732 (79.0)	21 (2.3)	41 (4.4)
8	1040	11 (1.1)	20 (1.9)	146 (14.0)	10 (1.0)	781 (75.1)	20 (1.9)	42 (5.0)
9	1020	12 (1.2)	16 (1.6)	146 (14.3)	16 (1.6)	768 (75.3)	17 (1.7)	45 (4.4)
10	971	7 (.7)	15 (1.5)	80 (8.2)	10 (1.0)	788 (81.2)	31 (3.2)	40 (4.1)
Total	9891	147 (1.5)	154 (1.6)	1033 (10.4)	196 (2.0)	7674 (77.3)	286 (2.9)	428 (4.3)

Tryout Test Design

There were 10 tryout forms. Each form required two testing sessions (Part 1 and Part 2). Table 5 below shows the tryout configuration for the HSPT in Communication Arts: Reading.

Table 5: Configuration of the Tryout

Item Distribution	Constructing/ Acquiring Meaning (Within the Reading Selections)	Metacognition (Knowledge about Reading)	Cross-Text/ Extending Meaning (Across the Reading Selections)	Composed Response/ Authentic Application (Response to the Reading Selections)
# of Multiple-Choice items	15	5	15	0
# of Extended-Response items	0	0	0	1

Each tryout test format included three or four reading selections presenting different points of view on the same common theme. Students were then asked to answer a series of multiple-choice questions. Twenty of the multiple-choice questions (Within the Reading Selections) required students to provide evidence of their understanding of the key concepts and ideas contained within each of the reading selections. Fifteen additional multiple-choice (Across the Reading Selections) questions required students to provide understanding of the key concepts and critical points common to two or more of the reading selections. Students were also asked to write one extended-response to the reading selections.

Five sets of testing materials were developed. Each set used the same reading selections, theme, and focus questions to produce two forms with different items. The purpose was to generate enough items out of two tryout forms for one operational form.

The reading tryout involved 9,891 students in grade 11 during the late fall of 1994. Each student took one tryout form. Since there were 10 forms and no items overlapped between any two forms, randomly equivalent group equating was used. To avoid exposing all forms in a participating school, forms were divided into four groups of triplets and two groups of quadruplets related by theme (Table 6). A school was randomly assigned to take only one group of forms. The forms within each triplet (or quadruplet) were then spiraled and administered to students within a classroom so that no students sitting next to each other would have the same form. This design permitted the equating of forms between triplets (or quadruplets) through the assumption of randomly equivalent groups of different participating schools taking the same form, but in different combinations. Forms in different triplets or quadruplets were equated by the Stocking and Lord (1983) procedure applied to the items in the common form. Additional information about equating will be presented in a later section of this report.

Table 6: Tryout Form Composition for Communication Arts: Reading

Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Form 1	Form 3	Form 5	Form 7	Form 10	Form 4
Form 2	Form 4	Form 6	Form 8	Form 1	Form 6
Form 3	Form 5	Form 7	Form 9	Form 2	Form 8
			Form 10		Form 9

Rating Process for Extended-Response Items

All multiple-choice items were machine-scored. All extended-response items were hand-scored by two readers. Readers were trained to implement the Michigan scoring guides. A number of quality control procedures were taken to ensure interrater reliability. Sets of actual student papers was used as anchor papers to illustrate responses exemplifying each of the possible score points for a response. Student responses were also used in check sets throughout the scoring process to ensure that readers were consistently applying Michigan standards. Table leaders conducted "read-behinds" by re-scoring sets of student responses to check the consistency of readers at their tables. Each extended-response question was worth 4 points. If the two readers disagreed by more than one point, a third reader was asked to adjudicate the scores. This situation rarely occurred. If two readings were sufficient, the item score was the sum of the two readings. If three readings were required, the item score was the sum of the three readings multiplied by 2/3, and rounded to the nearest integer. This process provided the extended-response items with 7 score levels in reading.

Interrater Reliability

Indices of interrater reliability, in the form of ranges of exact agreement and consistency, are presented by form in Table 7 on the next page. For this analysis, the agreement calculated for each reader is defined as the percent of times that the first reader agreed, within one point, with the second reader on the common items read by both readers:

$$\text{Agreement} = \frac{\text{\# of Items Reader 1 within One Point of Reader 2}}{\text{\# of Common Items Read by Readers 1 and 2}} \times 100 \qquad (1)$$

The agreement range describes the lowest and highest agreement rates seen among all readers. Consistency is defined as the percent of times the first reader agreed, within one point, with the second *or* third reader:

$$\text{Consistency} = \frac{\text{\# of Items Reader 1 within One Point of Reader 2 or Reader 3}}{\text{\# of Common Items Read by Readers 1 and 2}} \times 100 \qquad (2)$$

The consistency range spans the smallest and largest consistency rates among all readers. Consistency rates must be at least as large as agreement rates.

Both agreement and consistency ranges were generally small for the HSPT in Communication Arts: Reading tryout, with upper bounds achieving 100%. No form in Reading had an agreement range with a lower bound that dipped below 92%.

Table 7. Interrater Agreement and Consistency Ranges for the Tryout

FORM NUMBER	AGREEMENT RANGE	CONSISTENCY RANGE
1	94 - 98%	95 - 99%
2	94 -99	95 - 99
3	94 - 98	96 - 99
4	92 - 100	95 - 100
5	96 - 100	96 - 100
6	95 - 99	96 - 100
7	96 - 98	97 - 100
8	94 - 98	95 - 99
9	97 - 99	97 - 100
10	96 - 100	97 - 100

Tryout Statistics and Analyses[3]

Item Difficulty

Ranges of item difficulty (p-values) and item test correlations are presented in Table 8 (Appendix B). Rather than presenting the full range, which usually is not very informative because of the occurrence of outliers, the statistics are presented for the center 80 percent of the items in each form. That is, the items were rank-ordered in terms of p-values, and the values tabled for items at the 10th and 90th percentiles. For example, if a test had 40 items, p-values for the 4th and 36th most difficult items would be tabled. These ranges of p-values indicate that there was a good spread of item difficulties. Although not presented in this table, other analyses indicated that the extended-response items tended to be among the more difficult items in each form.

The "Collapsed Levels" columns in Table 8 indicate items where there were too few examinees who scored in a particular level so that scaling of that level for that item could not take place. In general, if there were fewer than 4 students with scores in a level, calibration could not occur. When calibration cannot occur, adjacent levels are collapsed. Although collapsing of levels can be important in a final operational calibration, collapsing of levels has little impact in a tryout. In reading, there were no items for which collapsing was necessary.

The average percentage of maximum score (%MS) ranged from 44% to 65% for the 10 tryout forms (Table 8). Thus, the test was moderately difficult for the student sample.

A final check was performed after the initial item analyses to identify items that were very difficult or had low item-test correlations. Four reading items were flagged for multiple correct answers; these items were not scored in any further analyses.

Tables 9 and 10 (Appendix B) present the raw means for each item and the item means at 5 quintiles. In general, the distributions of p-values spread relatively evenly within a form, with more items on the middle to higher end than on the lower end. While this implies that the items were fairly well distributed for this tryout sample, very few items had p-values below .20 or above .85. The means for the extended-response items ranged from 1.56 to 2.88 out of a maximum of six possible points.

[3] See Appendix B for Tryout Statistics.

Test Reliability

The reliability of a test indicates how well the test items "hang together." For the HSPT, reliability values are determined using internal consistency formulas, which indicate that the tests are measuring the same thing (within a particular test), and that students are answering consistently. Cronbach's alpha is used when there is a combination of multiple-choice and extended-response items.

The coefficient alpha reliabilities were reasonable for the number of items in the reading tryout, ranging from .73 to .88 (Table 8). Coefficient alphas were computed two ways, both including all items and excluding each individual item in each form of the HSPT tryout. The two outcomes were not statistically significantly different.

Content Validity

The current assessment is based on the *Michigan Essential Goals and Objectives for Reading Education*, which was approved by the State Board of Education in 1986. Because the current test is an achievement test used to endorse individual diplomas in reading, the most important type of validity to assess is content validity. To verify content validity, the test items must match the specified objectives given in the test blueprint or assessment framework.

Like all published achievement tests, the High School Proficiency Test in Communication Arts: Reading has a blueprint which indicates the objectives to be tested. Not all objectives are tested in any given form of a test. Both easy and hard items are used in every form of the test to balance the difficulty level of the items and to equate the different versions of the test to one another. The sample of items chosen for a version of the test represents the domain of all possible test items that fit the blueprint. For a student to do well on the test, he or she must have mastered the entire domain, not just bits and pieces.

As stated earlier in this report, the EDT in Communication Arts: Reading wrote all the tryout items based on the reading blueprint and framework documents. The CAC verified that each test question meets the objective it is supposed to measure, and fits the blueprint or framework. The BRC verified that the items are not disadvantaging any particular group.

Calibration Models

According to item response theory, item parameters are relatively invariant to changes in examinee groups. The important practical impact of this property is that the parameters of large numbers of items can be estimated even though each item is not answered by every examinee. This is known as person-free item calibration. The purpose of calibration is to estimate item parameters (e.g., item difficulty) as accurately as possible.

There are many calibration models. For the development of the HSPT, all calibration analyses were replicated using two sets of models, as recommended by the Technical Advisory Committee: (1) a combination of three-parameter logistic and two-parameter partial-credit models (3PL/2PPC) and (2) a combination of Rasch logistic and Rasch partial-credit models (1PL/PPC). The logistic models were used to analyze multiple-choice items and the partial-credit models were used to analyze extended-response items. The purpose was to compare which set would more appropriately reflect the data.

3PL/2PPC Model

The three-parameter logistic (3PL) model (Lord, 1980) allows items to vary in difficulty and discrimination and non-zero lower asymptotes ("guessing values"). It is commonly applied to multiple-choice items in tests like the HSPT, where guessing of correct answers can occur.

$$P_j(\theta) = P(X_j = 1 \mid \theta) = c_j + \frac{1 - c_j}{1 + \exp[-1.7a_j(\theta - b)]} \tag{3}$$

where θ = examinee's latent trait
a_j = item discrimination parameter for item j
b_j = difficulty parameter for item j
c_j = guessing parameter for item j
X_j = observed score for item j
$P_j(\theta)$ = probability of answering item j correctly given person ability θ

For the jth extended-response item with m_j levels, the item scores were integers ranging from 0 to $m_j - 1$. A two-parameter partial-credit (2PPC) model allows items to vary in both difficulty and discrimination. It was used to calibrate extended-response items (Yen, 1993). This model can be seen as a special case of Bock's (1972) nominal model and is the same as Muraki's (1992) "generalized partial credit model." The probability of a student with ability θ having a score at the kth level of the jth item is

$$P_{jk}(\theta) = P(X_j = k - 1 \mid \theta) = \frac{\exp(z_{jk})}{\sum_{i=1}^{m_j} \exp(z_{ji})}, \qquad k = 1, \ldots, m_j \tag{4}$$

where

$$z_{jk} = \alpha_j(k-1)\theta - \sum_{i=0}^{k-1} \sigma_{ji} \qquad i = 1, \ldots, k, \ldots, m_j \tag{5}$$

and
$\sigma_{j0} \equiv 0$.

α_j is the item discrimination. σ_{ji} is related to the difficulty of the item levels: the trace lines for adjacent scores levels intersect at σ_{ji} / α_j.

The 2PPC model is as follows:

$$P_j(\theta) = P(X_j = 1 \mid \theta) = \frac{1}{1 + \exp[-1.7a_j(\theta - b_j)]}, \tag{6}$$

Then,

$$a_j = \alpha_j / 1.7, \tag{7}$$

$$b_j = \sigma_{j1} / \alpha_j; \tag{8}$$

Conversely,

$\alpha_j = 1.7a_j$ and $\sigma_{j1} = 1.7a_j b_j$.

Rasch Model

The Rasch logistic model was used for multiple-choice items. This model allows items to vary in terms of difficulty, but all items were assumed to have the same discrimination (1.0) and a zero asymptote:

$$P_j(\theta) = P(X_j = 1 | \theta) = \frac{1}{1 + \exp[b_j - \theta]}. \tag{9}$$

Because of these simplified assumptions, for a two-level item,
$a_j = \alpha_j = 1$,

$b_j = \sigma_{j1}$.

Masters' (1982) Partial Credit model was used for the extended-response items. In formula,

$$P_{njx} = \frac{\exp \sum_{i=0}^{x} (\theta_n - b_{ji})}{\sum_{k=0}^{m} \exp \sum_{i=0}^{k} (\theta_n - b_{ji})}, \quad x = 0, 1, 2, \ldots, m_j \tag{10}$$

where P_{njx} is the probability of person n scoring x on extended-response item j.

Calibration Analyses

Item parameters and θ estimation were conducted using both the CTB-owned program PARDUX (Burket, 1991; 1995) and the commercial software BIGSTEPS (Linacre & Wright, 1993). PARDUX employs a marginal maximum likelihood procedure, implemented with an EM algorithm. Evaluations of the accuracy of the program with real and simulated data (Fitzpatrick, 1994) have found it to be at least as accurate as the Rasch program BIGSTEPS. The MEAP office traditionally uses BIGSTEPS.

For comparison purposes, BIGSTEPS estimates using the Rasch model were obtained in addition to the PARDUX analyses on the Communication Arts: Reading Form 1 in Group 1 and Mathematics Form 14 in Group 6. The correlations between parameters obtained by the two programs were 1.00. In summary, the two programs produced very similar estimates, with the estimates being the most similar for the item score levels where the most data were available.

Fit Statistics and Analyses

Item fit was evaluated with PARDUX by a statistic comparing observed and predicted trace lines. This fit statistic is a generalization of the Q_1 statistic described by Yen (1981).

Standardized fit values, referred to as Z statistics, can be compared over items and models. In addition, observed and predicted trace lines were compared graphically.

Rules of thumb were developed for flagging items for misfit. Recall that each item was scaled in two different samples. An item was flagged if it met either of the following criteria:

(1) Zs $\geq$ 4.0 in both samples, or

(2) (one Z $\geq$ 4.0) and (4.0 > the other Z $\geq$ 3.0), and a plot of expected and observed trace lines failed to demonstrate reasonable fit. (Note: A Z score is a standardized item fit score with a mean of zero and a standard deviation of 1.)

These rules of thumb for flagging misfit items can be compared in terms of stringency to the criterion used by CTB/McGraw-Hill for the tryout of multiple-choice items for major achievement batteries, such as the California Achievement Tests, and the Comprehensive Tests of Basic Skills. For those tests, Zs of 4.6 are flagged, even though their sample sizes are usually at least twice the size of ones used in the present study. As sample size increases, the power of the fit statistic increases. Thus, the flagging criteria used in this study is less stringent than the criteria used by CTB/McGraw-Hill in some other testing programs.

Summaries of item fit results are presented in Tables 11, 12 and 13. Far more items from the Rasch model had large Z values and were flagged for misfit than those from the 3PL/2PPC model. However, for the 3PL/2PPC model, there were items whose parameters could not be estimated, called non-convergent items. These items were often difficult items with low discrimination values. For the Rasch model, on the other hand, parameter estimates were convergent for all items. Thus, neither model effectively described an item performance when its observed trace line was essentially flat and had weak relationship to the predicted trace line. It should be noted that all the results shown here are from the software program PARDUX. Verification of the results from the software BIGSTEPS, which was designed specifically for Rasch model analysis, showed that some items that were misfit with the PARDUX were proved to be fit with BIGSTEPS.

Item Discriminations

The item discriminations were systematically lower for the extended-response items than for the multiple-choice items. On the average, the extended-response items had discriminations that were 25% of the values for the multiple-choice items for reading. Discriminations reflect how sharply performance can be categorized into successive score levels. It is not surprising that this categorization is less distinct with items that involved human evaluations of multiple levels of complex student performance.

The fact that the extended-response items had lower discriminations does not mean that these items are "less important" or contribute less information to the overall test score. The formula for item information is the following:

$$I(X_j \mid \theta) = a_j^2 \sigma^2 (X_j \mid \theta) \qquad (11)$$

The item information is a function of both the item discrimination (a_j^2) and the variance (σ^2) of the item scores. Items with more score levels tend to have substantially greater score variances, thus adding to the information they provide. Despite their lower discriminations, the extended-response items provided substantial amounts of information. Under the Rasch model, where all items are assumed to have the same discrimination, items with more score levels must be described as providing more information.

Table 14 (Appendix B) presents means and standard deviations of discrimination parameter estimates for all forms.

Equating

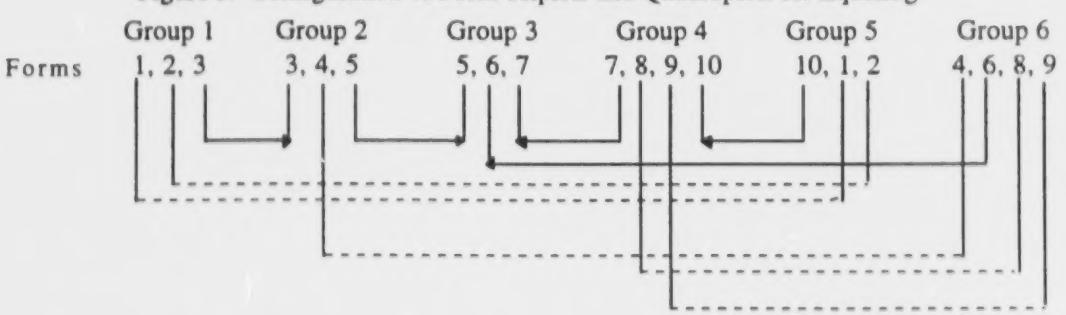
Test equating is necessary whenever one of two situations below occurs:

1. The tests are at comparable levels of difficulty and the ability distributions of the examinees taking the tests are similar. This is called "horizontal equating."
2. The tests are at different levels of difficulty and the ability distributions of the examinees are different. This is called "vertical equating."

For HSPT tryouts and pilots, horizontal equating was used because multiple forms were developed for each subject area and administered to randomly equivalant groups in the sample. The purpose of equating is to transform the scores of examinees taking form *X* to equivalent scores in form *Y* so that these scores can be compared to the scores of examinees taking form *Y*.

The equating process was conducted for both the Rasch and the 3PL/2PPC models here. The within-triplet theta (or scale score) distributions were aligned. The Stocking and Lord (1983) procedure was applied to the forms in common to the triplets (Forms 3, 5, 6, 7 and 10), as indicated by the solid lines in the following figure.

Figure 3. Configuration of Form Triplets and Quadruplets for Equating



The dotted lines indicate forms that were not included in the Stocking and Lord links (Forms 1, 2, 4, 8 and 9). These forms, therefore, could be used as a check on the adequacy of the equating. Forms 1 and 2 were of particular importance because the parameters from groups 1 and 5 were the "furthest apart" in terms of the linkings; that is, four Stocking and Lord links and five equivalent group links tied them together. By comparing the Form 1 test characteristic function based on the parameters from Group 1 to that based on Group 5, the adequacy of the link network could be double-checked. Similar checks could be done for forms 2, 4, 8 and 9. The checks showed that both models produced good equating results.

Scaling Model Selection

The advantages of using a Rasch model are its simplicity and elegance. Also, if data are scarce, Rasch model predictions tend to be more stable than those from a model with more parameters. The disadvantage of the Rasch model is that its simplifying assumptions may be inappropriate for a particular data set. The major advantage of the 3PL model is its less restrictive assumptions that permit more accurate description of data. The major disadvantage of the model

is that it requires a large number of examinees to provide sufficient data for parameter estimation. However, this is not a problem for the HSPT tryout.

For the HSPT tryout data, the Rasch models provided more misfit items (particularly for constructed-response items) than the 3 PL/2PPC models did, but the Rasch models did provide parameter estimates for all items. The 3PL/2PPC models produced better item estimates for most items but failed to converge for some other items in calibration (i.e., no estimates for those items). The TAC recommended the use of Rasch models over the 3PL/2PPC models for a large-scale assessment such as the HSPT, based on the empirical evidence and other technical considerations.

Racial and Gender Bias Analyses

Mantel Statistic for Ordered Response Categories

A Mantel-Haenszel methodology was used in the evaluation of the tryout items for differential item functioning (DIF). A statistic proposed by Mantel (1963) was obtained for specified racial and gender groups:

$$\chi^2 = \left[\sum F_k - \sum E(F_k)\right]^2 / \sum Var(F_k), \tag{12}$$

where F_k, the sum of scores for the focus group at the kth level of the matching variable, is:

$$F_k = \sum y_i n_{Fik}, \tag{13}$$

Readers are referred to Zwick et al. (1993) for a description of the terms of the statistic. The Mantel statistic, while necessary for the assessment of DIF in the extended-response items in each of the three content areas, reduces to the Mantel-Haenszel chi-square statistic (without continuity correction) when applied to the multiple-choice items. The Mantel statistic explicitly takes into account the possible ordering of the categories of the polytomous items, as opposed to a procedure proposed by Mantel and Haenszel (1959) that provides for a comparison of the reference and focus groups with respect to their entire response distributions. The Mantel statistic has a chi-square distribution with one degree of freedom.

Because the number of students in the minority groups taking each form was relatively small (almost always less than 200 per form), and the number of levels for the extended-response items was large (greater than five) when item scores were obtained by summing judges' ratings, the number of levels for the extended-response items was collapsed. After collapsing adjacent levels, the number of remaining levels that were evaluated for each extended-response item was half the maximum number of points plus one, or the same number of levels specified by the scoring rubrics for each item for each individual reader.

As specified by MDE for a sample of schools that were supplied to CTB/McGraw-Hill, item responses were analyzed for gender bias by evaluating DIF against females (focus group), with males as the reference group. The number of females in these analyses was large, approximately half of the roughly 1000 students who took each form.

The particular racial groups that were evaluated in the racial bias analyses were determined by the numbers of students in these groups who took the 29 tryout forms in the three content areas. The only group, excluding whites, that had appreciable numbers taking each form was African-Americans. Seventeen of the forms were administered to more than 100 African-Americans. The 12 forms that had fewer than 100 African-Americans were due to two schools with large

African-American enrollments dropping out of the sample and the failure to receive scores from a third school. A fourth school did not have as large an African-American population as expected.

After African-Americans, no defined racial group had consistently as many as 30 students taking each form. Consequently, Mantel statistics were obtained for a single (focus) racial group, African-Americans, treating whites as the reference group in the racial bias analysis.

Mantel racial and gender statistics were obtained for each form of reading by stratifying on total score. A total of 23 out of 360 reading items had a Mantel statistic that indicated racial DIF at a .05 significance level compared to 74 items that were flagged at the same significance level for gender DIF. The computation of standardized mean differences was employed to provide further estimation of item bias.

Standardized Mean Difference

Although the number of items that had significant Mantel gender statistics in each of the three content areas is substantially larger than the number of items having significant racial statistics, there are three reasons why the number of significant statistics cannot be considered to reflect the magnitude of DIF within each content area. First, the Mantel statistic is *asymptotically* distributed as chi-square, requiring a minimum expected number of five students within each of the cells defined by the combinations of strata and item levels. For the racial analysis, this assumption is frequently violated.

Second, a significant Mantel statistic rejects the null hypothesis of no DIF against the alternative hypothesis of DIF either against the focus *or* the reference group. Hence the number of significant Mantel statistics does not reflect solely DIF against the assessed focus group.

Finally, the much larger sample sizes for the female focus group relative to the African-American focus group results in more statistically powerful tests (i.e., tests that are more capable of correctly rejecting the null hypothesis of no DIF) in the gender analysis. The Mantel statistics for gender can detect the presence of smaller, and perhaps practically insignificant, amounts of DIF than the corresponding statistics from the racial analysis. An analysis of DIF that is more suitable to demarcating practically significant amounts of DIF across both racial and gender analyses would utilize an effect size index.

Unfortunately, while an effect size index in the form of the Mantel-Haenszel common odds ratio estimate, alpha, is available for the dichotomously scored items, no *single* analogous odds ratio-estimate is available for the polytomous items. However, the standardized mean difference (SMD) noted by Zwick et al (1993) offers an acceptable alternative:

$$SMD = \Sigma p_{Fk} m_{Fk} - \Sigma p_{FK} m_{RK}, \qquad (14)$$

where $p_{Fk} = n_{F+K}/n_{F++}$ is the proportion of focus group members who are at the *k*th level of the matching variable, $m_{FK} = (1/n_{F+K})\ (\Sigma y_t n_{RtK})$ is the mean item score for the focus group at the *k*th level, and $m_{RK} = (1/n_{R+K})\ (\Sigma y_t n_{RtK})$ is the analogous value for the reference group. As an effect size index, the SMD statistic takes into account the natural ordering of the response levels of the items and has the desirable property of being based only on those ability levels where members of the focus group are present. A positive value for a SMD reflects DIF in favor of the focus group.

Distributions of Standardized Mean Differences

Both racial and gender SMDs were obtained for the items in every form and are presented with the Mantel statistics. Ranges of the racial and gender SMDs for reading are:

Table 15. Ranges of Racial and Gender SMDs in the Tryout

Content Area	Racial	Gender
Communication Arts: Reading	- .32 to .21	- .15 to .40

An evaluation of both the Mantel and the SMD statistics for the racial comparisons suggested that levels of standardized mean differences that have practical significance could be determined. Statistically significant ($p = .05$) racial Mantel statistics were often associated with SMDs that had absolute values of .10 and greater. Setting a criterion of -.10 for a determination of practically significant DIF, representing a one tenth of a score point decrement in focus group performance relative to the reference group (controlling for ability), would allow a goal of limiting the conditional between-focus-and-reference-group difference to no more than one score point in any form. The distribution of SMDs for reading below appears to permit the construction of forms having 10 or fewer items demonstrating DIF either against a racial or against a gender group that an individual form could have and still attain the maximum one-score-point-conditional-group-difference goal. A maximum of one score point difference is desirable, given the high-stakes nature of the test.

Table 16. Frequency Distribution of Tryout Items by Racial SMDs

(SMD≤-.30)	(SMD≤-.20)	(-.19≤SMD≤-.10)	(-.09≤SMD≤.09)	(.10≤SMD≤.19)	(SMD≥.20)	(SMD≥.30)
1 items	2 items	17 items	320 items	20 items	1 item	0 items

Table 17. Frequency Distribution of Tryout Items by Gender SMDs

(SMD≤-.30)	(SMD≤-.20)	(-.19≤SMD≤-.10)	(-.09≤SMD≤.09)	(.10≤SMD≤.19)	(SMD≥.20)	(SMD≥.30)
0 items	0 items	7 items	339 items	7 items	7 items	4 items

Large numbers of reading items that were tested in a pair of forms afforded the possibility of evaluating the stability of the designation of a practically significant SMD of -.10 or less. Because only DIF against the focus group is being assessed, SMDs greater than .10 are considered as not representing DIF. The racial and gender SMDs for the 83 reading items that were in a pair of forms were compared across form pairs. A relatively small number of these items, four (5% of the 83) had a practically significant racial SMD for one form but not the other. Five items (6%) had one practically significant and one not practically significant gender SMD. This results in a minimum of 95% of the items having common designations of not practically significant racial DIF and 94% of the items having a common designation of not practically significant gender DIF. Hence, a value of - .10 as a criterion of practical significance appears promising in producing relatively stable classifications of DIF over item administrations.

Overall DIF Rating

The distribution of racial and gender SMDs under the criterion of -.10 for practically significant DIF allows the construction of an overall rating of DIF that combines both racial and gender

DIF against the focus groups. An overall rating is a useful index in the development of the pilot or operational forms. Content editors can utilize test development software to select items in a manner that minimizes DIF against both focus groups.

A useful overall index of DIF might allow several gradations of the practical severity of both racial and gender DIF. An item could be considered to manifest a lower degree of practically significant DIF against a racial or gender group if the SMD ranged between -.10 and -.19 and a more serious degree of DIF if the SMD was less than or equal to -.20. An item would accumulate one point on the overall rating scale if the racial SMD fell in the former category and two points if the racial SMD fell in the latter category. Similarly, an item would accumulate an additional point on the overall scale if the gender SMD fell in the former category and two points for the latter. Consequently, if an item demonstrates neither of the two levels of practically significant racial DIF and neither of the two levels of practically significant gender DIF, the item's overall rating would be one (zero would seem to be a less desirable alternative because it connotes the absence of DIF). An item would obtain the maximum overall rating of five if both racial and gender DIF was of the more serious kind. An overall rating of two would imply the item had a racial or gender SMD between -.10 and -.19, but not both. An overall two, three, or four could be obtained by various combinations of lower and higher levels of practically significant racial and gender DIF. All possible overall ratings are described in the table below.

Table 18. Overall DIF Rating Classification as a Function of Gender and Race

	Race DIF		
Gender DIF	(.09 ≥ SMD ≥ -.09)	(-.10 ≥ SMD ≥ -.19)	(-.20 ≥ SMD)
(.09 ≥ SMD ≥ -.09)	1	2	3
(-.10 ≥ SMD ≥ -.19)	2	3	4
(-.20 ≥ SMD)	3	4	5

Table 19. Frequency Distribution of Items by Overall DIF Rating

DIF Rating	1	2	3	4	5
# of items	335	22	3	0	0

Detailed DIF statistics are presented in Table 20 (Appendix B).

Items with a DIF rating of two or higher were subject to an additional review by the Bias Review Committee and the Content Review Committee for any apparent bias. If none was found and the item was determined to adequately measure the test content, it was kept.

Pilot Test

Items that survive the tryout stage are then piloted before they are used in an operational test. Frequently, 25-50% of items tried out are discarded at the tryout stage. Based on review of the HSPT in Communication Arts: Reading tryout results, CTB worked with the CAC and MDE staff to refine items and scoring rubrics before piloting began. The HSPT in Communication Arts: Reading started out with five sets of reading selections which were used to generate twice the number of test items needed for the operational test forms. The rationale for developing an overage of test items of this quantity was based on MEAP's prior experience of attrition of

reading test items after tryouts. Doing so provided enough items to produce two tryout forms per set of reading selections for a total of ten tryout forms.
Subsequent to the tryout, the Content Advisory Committee recommended not piloting two sets of reading materials because insufficient numbers of test survived. Consequently, only three forms were piloted based on the CAC recommendation, with one set of reading selections for each form.

The purposes of the HSPT in Communication Arts: Reading pilot administration were to:

- check if revisions based on the tryout were successful, or whether an item should never be used;
- produce 3 equivalent forms of the High School Proficiency Test in Communication Arts: Reading that could be used interchangeably in future administrations;
- examine characteristics of the revised items in each form; and,
- examine technical soundness of the reconstituted forms for operational administrations.

CTB made all necessary revisions of the assessment materials suggested by the CAC and MDE. They also prepared the test booklets, answer documents, administrator's manuals and all supporting materials for the pilot administration.

Pilot Sampling

As in the tryout, the target population for the pilot was all eleventh graders in Michigan, including students in both public and private schools. The sampling procedure was also the same. Fewer schools were sampled in the pilot because fewer forms were tested. However, the proportions of participating students by gender and ethnicity were very similar to that of the tryout. When a sampled school declined to participate in the pilot, a substitute school with similar characteristics was replaced to keep the representativeness of the sample. The number of students taking each form is listed in Table 21 below.

Table 21. Number of Students Participating in the Pilot by Form

Form	# of Students
1	1505
2	1448
3	1396
Total	**4349**

Pilot Administration

Sample schools were asked to test all eleventh grade students during a five-day administration window in April 1995. Classroom teachers were asked to administer the test. For security purposes and to minimize the exposure of test forms, makeup testing for students who were absent during the pilot was not recommended.

General Results

Table 22 (Appendix C) provides a summary of the descriptive statistics for both the complete sample that took a form and the two constituent subsamples taking the same form as it was administered within spiraled sets of two forms. Complete sample form means for the eight reading forms ranged between 21.22 (Form 3) and 23.60 (Form 1) out of 38 possible points.

The mean p-values were between .56 and .62 for the test forms. This indicates that the pilot test items were moderately difficult for the 11th grade student sample. Considering each form as a whole, the mean item-test correlations were in the .40s and the coefficient alphas were in the .80s for all forms. Both of these statistics were fairly high, implying that the items were very consistent internally.

Interrater Agreement

Scorers were hired and trained by CTB to score the extended-response items for the pilot test, using Michigan standards. On the pilot, the one extended-response item in each form was worth four points. Scores for extended-response items were obtained by averaging the ratings of two or three judges and rounding to the nearest integer. Only when the two readers' scores were not the same or adjacent - that is, more than one point apart on the same item - was the third reader introduced. Table 23 contains ranges for readers agreement and consistency for the pilot forms. Excluding one index computed for a reader who read only three papers (indicated in [ ] in Table 23 for Form 3), consistency indices ranged between 90% and 100%.

Table 23. Interrater Agreement Ranges

FORM NUMBER	AGREEMENT RANGE (%)	CONSISTENCY RANGE (%)
1	86-100	90-100
2	89-100	90-100
3	91-100 [67(3)]*	93-100

Agreement: percentage of times that a reader agreed, within one point, with the second reader.
Consistency: percentage of times that a reader agreed, within one point, with the second or third reader.
*One reader completed only three readings for Form 3 with an agreement rate of 67%. The next lowest agreement rate for this form was 91%.

Table 24. Mean Interrater Agreement Between The First Two Readers

Pilot Forms	Rates of Agreement		
	Exact Agreement	Adjacent Agreement	Non-adjacent Scores
Form 1	68.0%	25.1%	6.9%
Form 2	67.5	26.2	6.3
Form 3	77.0	18.5	4.5
All Forms	70.7	23.4	5.9

The mean rate of exact agreement was at least 68% for all items (Table 24). The rate of non-adjacent reader scores was less than 7%. It should be noted that each extended-response item had from 201 to 260 students choosing to leave the item blank. More detailed interrater reliability statistics are presented in Table 25 of Appendix C.

Different approaches to weighting the multiple-choice and extended-response items were discussed at a TAC meeting. It was decided that all items would keep the raw score weight.

Group Descriptive Analyses

Descriptive statistics for four groups--whites, African-Americans, females, and males--are presented for each of the three reading forms in Table 26. On all forms of reading, females consistently have higher mean scores than males and white means are higher than African-American means. It is important to note that African-American means in Table 26 are based on less than 200 students for all reading forms. The relatively small number of African-Americans may be attributed to the difficulty of getting a large number of high schools in metropolitan and other urban areas with large African-American enrollments to participate in the pilot. The difference in group means was generally smaller for the three reading forms than for the mathematics or science forms.

Table 26. Michigan HSPT in Communication Arts: Reading Pilot
Group Descriptive Statistics

Form	White			African-American			Female			Male		
#	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
1	23.97	7.06	1151	21.43	6.27	173	24.70	6.38	769	22.42	7.46	724
2	22.84	7.10	1127	17.62	6.07	164	23.09	6.80	726	20.78	7.49	709
3	21.90	7.53	1048	17.90	6.92	190	22.66	7.06	698	19.82	7.91	686

Gender/Ethnicity DIF Statistics

Table 27 (Appendix C) contains DIF (differential item functioning) statistics, in the form of standardized mean differences (SMDs) for two group comparisons: males versus females and whites versus African-Americans. The SMDs for each comparison were partitioned into four groups in accordance with the procedure used for the tryout forms.

The three reading pilot forms were extended using the tryout DIF statistics, to ensure that the absolute difference in the amount of DIF for whites versus African-Americans **and** the absolute difference in the amount of DIF against males versus females was no greater than three. The purpose of constraining the absolute difference in DIF to no more than three for each of the two group comparisons was to ensure that DIF was relatively balanced across each of the two groups in each of the two comparisons.

The absolute difference in the amount of total DIF for the 6 comparisons (2 comparisons times 3 forms) can be seen in Table 27, within each pair of evaluated groups. The differences were frequently very small, with no comparison exceeding an absolute difference of 3.

Summary

In summary, even though the test was moderately difficult, all the pilot forms showed high test reliability. Students had more difficulty answering extended-response items than multiple-choice items. In fact, a fairly high proportion of students did not respond to the extended-response items. The interrater agreement between the two readers for the extended-response items was relatively high.

Part 4. Student Survey and Teacher Survey

The Technical Advisory Committee (TAC) recommended that a study be done prior to the first administration of the Michigan High School Proficiency Test and again just prior to the time when the first graduating class would be impacted.

In early 1994, planning for an opportunity to learn study began. It was tentatively agreed that the final responsibility for the design must reside at the State Department level, that members of the Framework Committees should be involved in the design, that teachers in every district needed to be surveyed, that students should be sampled, and that the TAC should review the sampling plan and the draft survey instrument(s).

In March 1994, one TAC member, Department staff, and a member of the Communication Arts: Reading Framework Committee reached two major decisions:

(1) Surveys would be sent to every high school to the subject matter coordinators for the content areas tested on the HSPT. They would be asked to form committees of teachers from their high schools as well as their feeder schools to fill out the survey.
(2) A sample set of students would be part of the study.

In subsequent meetings with the Communication Arts: Reading Framework Committee, discussions were held regarding the content and the format of the surveys. It was agreed that the general form of the surveys was to be the same across content areas, but that form should not take precedence over substance and if there were good reasons for having different formats, it would be allowed. Content area experts were to be responsible for the actual wording of the surveys.

The study was originally intended to address three purposes: (1) to help make adjustments to the tests if necessary, (2) to aid in standard setting and (3) to provide schools with information that could be used for professional development.

On September 2, 1994, an overview of the proposed design was presented to the TAC. The TAC members suggested that the names of the surveys be changed from "opportunity to learn" surveys to the "Teacher Survey" and the "Student Survey." Revisions were suggested and made for the Student Survey. The Teacher Survey was discussed at length, reviewed and revised. Both the student and teacher surveys were piloted at several sites before being sent out.

Communication Arts: Reading Student Survey Results

The Communication Arts: Reading Student Survey (see Appendix D) was given to the students who participated in the reading tryout. The students completed the survey prior to taking the item tryout "tests" so that student perceptions pertaining to performance would not influence survey responses.

The Communication Arts: Reading Student Survey contained 23 statements. The common stem was as follows: "By the end of tenth grade, how often did your school experience include:..." Students were to respond on a four-point scale from "never" to "a lot." Note that "never" was translated to a value of "zero" (0), "very little" to "1," "some" to "2," and "a lot" to "3."

Table 28 below presents the summary data for the student survey results. The mean score for the 23 reading survey questions was 1.71 (2.00 = some). The lowest mean for a survey

question was .79, which places it between "never" and "very little." Only five questions (22%) had a majority of the students respond less than "some." Five questions (22%) had a mean less than 1.5.

The reading test tryout was made up of questions of four types: constructing meaning, cross/text, knowledge about reading, and extended-response. There were survey questions relating to these four constructs of reading, as well as some general questions. A few questions related to more than one reading construct. By part, the mean survey scores ranged from a low of 1.53 for "General" to a high of 2.01 for "Construct Meaning."

Because the surveys were given to the same students who participated in the tryout, it was possible to correlate the mean scores for the students on the survey with their scores on the tryout tests. The correlations are positive, but not particularly high (.2028). Thus, the students' perceptions of whether they were taught something did not seem very highly related to how they actually scored on the tryouts.

Among the content areas, it appears that the student survey results in reading were not as positive as mathematics but better than science. However, the reading survey had the lowest overall mean (1.71).

Table 28: Student Survey Results Summary
Content: Communication Arts: Reading

Total # of questions	23
overall mean	1.71
lowest mean	.79
# & % of questions that the majority marked less than "some" (2.0)	5 (22%)
# & % of questions with a mean less than 1.5	5 (22%)
correlation statistic of survey mean and tryout score	.20

Conclusions From The Student Survey

In drawing conclusions from the student survey results, one must keep in mind that there was no good way of determining how honestly students responded to the questions or even the extent to which they understood the questions. Given those cautions, it was concluded that school experiences in general included the types of activities useful in assisting students to learn the content to be tested on the proficiency test.

Communication Arts: Reading Teacher Survey

The Communication Arts Teacher Survey was sent to English department chairpersons and reading supervisors at all high schools in the state (N=758), May of 1995. Each was asked to form a team of teachers to work with them in completing the Teacher Survey and an Instructional/Curriculum Support Materials Form (which they did not need to return).

The Communication Arts teacher survey is composed of 50 statements (24 writing and 26 reading) organized by parts and objectives within a part. For reading, the two parts are "Types of Reading" and "Objectives." For "types of reading", respondents completed only the column that directed them to circle all grades receiving instruction (Column I). For the objectives part, respondents indicated all grades receiving instruction and were asked to circle the one grade at which sufficient classroom instruction had occurred to expect understanding/proficiency (Column II).

Summary Of The Teacher Survey Results

In summarizing the Communication Arts Teacher Survey results, it must be remembered that the data analyzed were based on a low return rate of 245 responses out of 758 surveys sent to schools. So, the responses may not be representative. Nevertheless, some tentative findings emerge from the teacher survey results that are summarized in Table 29:

- only one of the 26 statements had more than 50% of the schools circle the "NT" (Not Taught) response;
- only two statements had more than 25% of the schools circle "NT";
- no statement had 50% or more of the schools circle "NSI" (Not Sufficient Instructions);
- one statement had fewer than 25% of the schools circle "NSI"; and
- five statements had "NSI" circled by fewer than 10% of the schools.

Table 29. Teacher Survey Results Summary
Content: Communication Arts: Reading

# and % of statements where NT circled by 25% or more	2 (8%)
# and % of statements where NSI circled by 50% or more	0 (0%)
# and % of statements where NSI circled by 25% or more	3 (21%)
# and % of statements where NSI circled by less than 10%	5 (25%)

Overall Summary And Follow-Up[4]

Both the student and teacher survey results suggested that many of the objectives were already being taught in the majority of the schools and that they were sufficiently taught for students to

[4] In July 1996, the State Board of Education approved the standards as set by the standard setting committees, without changes. Information about the student and teacher surveys is adapted from a paper presented at the 1996 Michigan School Testing Conference by Mehrens, Smolen and Yan. Ann Arbor, MI.

have proficiency in them. However, in reading, there were a few objectives that were judged to have not been taught with sufficient thoroughness. Additional survey information is provided in Tables 30-33 in Appendix D.

The results of both the teacher and student surveys were presented to the standard setting committees at the time they made recommendations regarding scores. Prior to that time, the department devoted considerable time determining just how the data should be presented and what the committees should be told about the relevance of the data for standard setting. It must be stressed that these data were gathered in the 1994-95 school year, and that information about the content of the proficiency tests continued to be widely disseminated before the test was given in the spring of 1996. It is reasonable to believe that instruction in the schools has become more aligned to the objectives tested as time has passed.

The results of these surveys were disseminated to curriculum coordinators in the schools who were encouraged to use them in planning curricular/instructional changes prior to the first administration of the HSPT. It should have been clearly understood by local schools that it is in the best interests of their students to teach them material from a content domain that is sampled on a test whose passing is a requirement for a state-endorsed certificate.

REFERENCES

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 37, 29-51.

Burket, G. R. (1991; 1995). PARDUX. Monterey, CA: CTB Macmillan/McGraw-Hill.

Fitzpatrick, A. R. (1994). Two studies comparing parameter estimates produced by PARDUX and BIGSTEPS. Unpublished manuscript.

Linacre, J. M. & Wright B. D. (1993). A Users Guide to Bigsteps: Rasch Model Computer Program. Chicago: MESA.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. Journal of the American Statistical Association, 58, 690-700.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of a disease. Journal of the National Cancer Institute, 22, 719-748.

Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-174.

Mehrens, W. A., Smolen, D. L. & Yan, J. W. (1996). Michigan High School Proficiency Test. Summary of Student and Preliminary Teacher Survey Results. Paper presented at the 1996 Michigan School Testing Conference, Ann Arbor, MI.

Michigan State Board of Education (1980). The Common Goals of Michigan Education. Lansing, MI.

Michigan State Board of Education (1986). Michigan Essential Goals and Objectives for Reading Education. Lansing, MI.

Michigan State Board of Education (October, 1991). Michigan Model Core Curriculum Outcomes. Lansing, MI.

Michigan State Board of Education (1994). The Assessment Framework for the Michigan High School Proficiency Test in Communication Arts: Reading. Lansing, MI.

Muraki, E. (1992). A generalization partial credit model: Application of an EM algorithm. Applied Psychological Measurement, 16, 159-176.

Stocking, M., & Lord, F. (1983). Developing a common metric in item response theory. Applied Psychological Measurement, 5, 245-262.

Yen, W. M. (1981). Using simulation results to choose a latent trait model. Applied Psychological Measurement, 5, 245-262.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. Journal of Educational Measurement, 30, 187-213.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. Journal of Educational Measurement, 30, 233-251.

Appendix A

Expert Panel*

Mr. Thomas Fisher
Administrator of Student Assessment Services Section
Florida Department of Education

Ms. Sharon Johnson-Lewis
Director of Planning, Research and Evaluation
Detroit Public Schools

Ms. Marjorie Mastie
Supervisor for Assessment Services
Washtenaw Intermediate School District

Dr. William Mehrens, Expert Panel Chair
Professor of Educational Measurement
Michigan State University

Dr. Jason Millman
Professor of Educational Measurement
Cornell University

Dr. Susan Phillips
Associate Professor of Educational Measurement
Michigan State University

Dr. Edward Roeber
Director of Student Assessment Programs
Council of Chief State School Officers

Dr. Roger Trent
Director, Division of Educational Services
Ohio Department of Education

* Job titles at time panel convened.

Technical Advisory Committee (TAC)*

Dr. Gail Baxter
Assistant Professor of Education
University of Michigan

Dr. Roger Trent
Director
Assessment and Evaluation
Ohio Department Of Education

Ms. Sharon Johnson-Lewis
Assistant Superintendent
Research, Development & Coordination
Detroit Public Schools

Dr. William Mehrens
Professor of Educational Measurement
Michigan State University

Dr. Edward Roeber
Director, Student Assessment Programs
Council of Chief State School Officers

Dr. Joseph Ryan
Research Consultant Center
Arizona State University West

* Job title at time of HSPT development

Exercise Development Team - Communication Arts: Reading (EDT)*

Marilyn Carlsen
Staff Development Trainer
Farmington Training Center
Farmington Public Schools

Alicia Coquitt
Reading Specialist
Curriculum Office
Detroit Public Schools

Joyce Ilges
Retired English Teacher
Troy School District

Marcia Klemp
Reading Teacher
Lakeshore Middle School
Grand Haven Area Public Schools

Douglas Luke
Special Education Instructor
Grand Rapids Public Schools

Marian Schultz
Reading Teacher
West Middle School
Rochester Community Schools

Patricia Holmes
English Department Head
Redford High School
Detroit Public Schools

Marilyn Whitlow
English Instructor
Muskegon High School
Muskegon Public Schools

Charlie Peters
Language Arts Consultant
Oakland Schools

Terri Terry
Media Specialist
Pinckney High School
Pinckney Community Schools

* Job title at time of HSPT development

Content Advisory Committee (CAC)* - Reading

Josephine Belford
Supervisor
Office of Communication Arts
Detroit Public Schools

Marc Conley
Professor of Content Reading
Michigan State University

Larry DeVoogd
Director
Compensatory Education
Muskegon Public Schools

Kimberly Hurtte
English Teacher
Renaissance High School
Detroit Public Schools

Beverly Kozin
Teacher
West Bloomfield High School

Charlie Peters
Secondary Reading Consultant
Oakland Schools

Karen Urbshchat
Consultant
Wayne County Regional Educational
Service Agency

Deanna Birdyshaw
Curriculum Resource Teacher
Ypsilanti Public Schools

Sheila Potter
English Language Arts Consultant
Michigan Department of Education

* Job title at time of HSPT development

Bias Review Committee (BRC)*

Mr. Aden D. Ramirez, Director
Bilingual/Migrant Program
West Ottawa Public Schools

Ms. Ellen Carter-Cooper
Educational Consultant/
School Development Unit
Michigan Department of Education

Ms. Stephanie Rockette
Mathematics Resource Teacher
Vincent Place/Teacher Resource
Benton Harbor

Dr. Elana Izraeli, District Coordinator
Testing & ESL Programs
West Bloomfield School District

Mr. H. William Leavell, Jr.
Research Specialist
Michigan Jobs Commission/Michigan
Rehabilitation Services

Mr. Jesus M. Solis
Educational Consultant
Michigan Department of Education

Mr. William Gay
Teacher/Huron High School
Ann Arbor School District

Dr. Pauline Coleman
English Language Arts Coordinator
Ann Arbor Public Schools

Mr. Robert Brown
Huron High School
Ann Arbor School District

Dr. Rossi Ray-Taylor
Director of State and Federal Programs
Lansing School District

Ms. Marian Phillips (replaced Dr. Ray-Taylor)
Supervisor, Research and Evaluations
Lansing School District

* Job title at time of HSPT development

Expert Panel Recommendations

1. The State Board should not specify subject areas other than Communications Skills, Mathematics, and Science for the initial assessment.

2. Communication skills assessed during the first assessment cycle should be limited to reading and writing.

3. The State Board and the Michigan Department of Education need to determine which subsets of the model core curriculum should be included in the assessments. This needs to be done very shortly. The decision should be based on recognition of the importance of students' opportunity to learn the content and some knowledge regarding what is likely to be in the school curricula by the date of the first test. The decision should not be that the total core curriculum is the appropriate domain from which to build the tests.

4. Once a determination is made regarding the testable portion of the core curriculum, there should be an administrative rule or statute that specifies this portion of the core is exempted from the permissive language in P.A. 25 and must be taught by the local districts to all students.

5. Once the testable portion of the core is determined, there should be wide publicity of this to the local districts. Consideration should be given to how this information can be disseminated with enough detail to let students and educators know the knowledge and skills to be tested but without so much detail that the students can answer the questions without understanding the curricular elements from which the items are only a sample.

6. Gather evidence from both teachers and students regarding the opportunity to learn the content domain the tests sample prior to the first administration.

7. Provide instructional support and training to local teachers if there is a need.

8. The State Board should not make any changes in the core curriculum or selected testable core prior to 1997.

9. When (or if) any changes are made in the core curriculum, there must be a phase-in period, and the tasks described in recommendations 3 through 7 would need to be repeated.

10. Name the assessment the "Michigan High School Graduation Tests."[5]

11. The Department of Education should caution its employees and the State Board against making any unsubstantiated statements about what the tests measure or what inferences can be made from the test scores. There should be an official statement about the tests and the inferences that can be drawn from the scores.

12. Demand that the test developer design sufficient safeguards to ensure that the test adequately samples the defined content.

13. Be careful not to make any official statements that would suggest the test has criterion-related validity if supportive data have not been gathered.

---

[5] Because there will be different tests for different content areas, we suggest the plural "tests." However, for ease in subsequent writing we will, at times, refer to the total assessment as a test. When we do so, it should be understood that the reference includes all the tests.

14. Contract for enough items initially so that after losses through pilot and field testing there will be enough to build forms through the 95-96 administration year.

15. Reissue a contract in sufficient time to have items developed and tried out (possibly embedded in a live form) prior to their being needed for the 96-97 year.

16. Schedule a large scale field tryout for tenth graders by the spring of 1994.

17. Appoint and train a standard-setting committee.

18. Use a technical advisory committee to help develop a specific standard-setting procedure.

19. The State Board of Education should establish a passing score through administrative rule based upon a recommendation by the superintendent of public instruction with the advice of appropriate committees.

20. Consider setting incremental cut scores for different graduating classes at the time the State Board of Education makes its initial decision.

21. The item sensitivity reviews should be completed by a committee that is selected and trained specifically for this task. Most members should represent Michigan's predominant minority groups. However, it would be wise to have at least one member of the committee be a minority group member from out-of-state who is a recognized expert in the area.

22. Statistical item bias studies should be conducted. Items which show up as statistically biased should be reviewed (but not necessarily discharged) by an item bias committee (conceivably, but not necessarily the committee used for the item sensitivity review) and a content review committee.

23. Obtain the following reliability estimates: internal consistency, interrater reliability, generalizability across writing samples, and the reliability or standard error at the cut score.

24. Scores should be reported as "Pass" or "Fail." Those individuals who fail should be given some information regarding how close they were to passing, and they should be given some diagnostic information that would facilitate remediation efforts. There are important technical details (e.g., reliability of difference scores) regarding various methods of reporting diagnostic information and specific plans should be formulated by a technical advisory committee prior to approval of the final test specifications.

25. We would encourage use of a common scale across subject matter areas. This takes some advance planning to avoid adopting a scale that is appropriate for one test, but unworkable for another.

26. Develop detailed rules (procedures) for designating forms for make-up examinations and out of school (i.e., Adult Ed.) populations. Determine whether you should ever reuse a form. Determine how many times you will administer the test each year. Determine equating procedures (e.g., number of anchor items to be used). Based on these considerations, initially develop enough alternate forms to last through at least the 1995-96 school year. Start developing more forms/items prior to that so a sufficient supply is continuously available.

27. Use a technical advisory committee to help develop specific equating procedures.

28. Consider carefully policies regarding all test administration conditions. For example, the decision of whether or not to use calculators in the mathematics test must be made by the department, not by local school personnel. Train local school personnel adequately to administer the tests. Consider random auditing of the administration process to ensure uniformity throughout the state.

29. Be cautious about any "predictive" interpretation of the scores of any single individual from testing in earlier grades. Such tests should be thought of as providing only an early awareness.

30. The department should prepare and have the board adopt written procedures regarding make-up examination provisions.

31. The department should prepare and have the board adopt specific written rules regarding the number of retakes that should be allowed, and how many attempts a student should be given prior to the time he/she is scheduled to graduate.

32. Develop a detailed proposal that addresses questions regarding remediation efforts and the respective responsibilities of the state, the district and the student for remediation efforts.

33. Enact an administrative rule regarding testing issues related to special education students and students with limited English proficiency.

34. Individuals in adult education programs who wish to receive high school diplomas after the end of the 1996-97 school year should be required to pass the High School Graduation Test.

35. Obtain the services of the Attorney General's Office early on in the process and continuously as new policies are developed and implemented.

36. The State Superintendent of Public Instruction and the State Board of Education should work with the legislature to adopt statutory authority for the high school graduation testing program.

37. Carefully investigate liability issues with assistance from the Attorney General's Office. Attempt to obtain necessary statutes with respect to liability. Inform all committees and all staff regarding their potential liability.

38. Schools should be notified immediately regarding this graduation requirement and the information disseminated to all teachers. Students and their parents should be notified no later then the spring of 1993.

39. The department should prepare, and the board should adopt, detailed policies regarding what should be documented and how long the documentation should be kept on file. We generally suggest that all documentation be kept for a period of at least five years following the school year in which the test was administered. We suggest keeping "forever" the initial development documentation and records about when, why, and how procedures are adopted and/or changed.

40. In consultation with the Attorney General's Office, and based in part upon discussions with representatives of state education associations (e.g., teachers' unions and administrators' associations), the department should prepare, and the State Board of Education should adopt, rules regarding what constitutes inappropriate behavior on the part of educators or students with respect to test-taking behavior, security issues, and so forth; and what

penalties will be imposed for violation of these rules. These rules and the penalties should be disseminated to educators and students prior to the initial administration of the graduation test.

41. The department needs to develop a complete list of rules/regulations that need to be adopted and decide whether these can simply be adopted by the board or whether they need legislative approval.

42. Detailed security arrangements need to be developed.

43. Detailed policies regarding security valuations need to be established. Staff should investigate current laws regarding freedom of information exclusions, and if they are insufficient, request new legislation to exempt secure test materials from the freedom of information regulations.

44. The department needs to determine what additional equipment/facilities are needed for storage of secure materials, shredding out-of-date secure materials, etc.

45. An annual test administration plan should be developed and disseminated to all school districts.

46. The tests should first be administered to 10th graders in the spring of 1995 and they should be administered at least twice each in the junior and senior years.

47. The department should conduct a careful study to assess additional staffing needs in assessment and instructional programs.

48. The position of supervisor of state assessment should be filled as quickly as possible.

49. The following advisory committees should be appointed: 1) a Michigan Department of Education Steering Committee, 2) a Testing Policy Advisory Committee, 3) a Bias Review Panel, 4) a Technical Advisory Committee, 5) a Content Review Committee in each content area of the test, 6) an overall content review committee, and 7) a Standard Setting Committee.

50. Use at most two contractors: one for test development and formal field tryouts; and another for test administration, scoring, and reporting.

51. Obtain more detailed information from other states with similar programs regarding fiscal needs. Make recommendations to the legislature that are sufficient to cover department needs, and make clear to them that the task simply cannot be done without adequate support.

Michigan School Stratum Classifications

The Michigan schools are classified into seven strata relative to populations where the schools reside.

1. Large City
   Central city of a Metropolitan Statistical Area (MSA) with a population greater than or equal to 400,000 or a population density greater than or equal to 6,000 people per square mile.

2. Mid-size City
   Central City of an MSA with a population less than 400,000 and a population density less than 6,000 people per square mile.

3. Urban Fringe of Large City
   Place within an MSA of a Large Central City and defined as urban by the Census Bureau.

4. Urban Fringe of Mid-size City
   Place within an MSA of a Mid-size Central City and defined as urban by the Census Bureau.

5. Large Town
   Town not within an MSA and with a population greater than or equal to 25,000 people.

6. Small Town
   Town not within an MSA and with a population less than 25,000 and greater than or equal to 2,500 people.

7. Rural
   A place with fewer than 2,500 people and coded rural by the Census Bureau.

Criteria for Writing and Editing Multiple-Choice Items

- ❒ The item is free of gender, ethnic, racial or other bias.
- ❒ The content of the item is grade-appropriate.
- ❒ The reading level of the item stem and answer choices is suitable for the student being tested.
- ❒ All factual information has been checked and documented against reliable, up-to-date sources.
- ❒ A student possessing the skill being tested can clearly select one and only one correct response.
- ❒ All extraneous material has been edited from the stem.
- ❒ All item distractors are plausible to someone who has not mastered the skill being measured.
- ❒ Answer choices are free of repetitious words or expressions that can be included in the stem.
- ❒ All answer choices are consistent with the stem both conceptually and grammatically as well as consistent with each other.
- ❒ All answer choices are mutually exclusive.
- ❒ All answer choices in the item are approximately equal in length (i.e., no one choice is much longer or shorter than another).
- ❒ No outliners - answer choices that are obviously different from the others.
- ❒ The correct response for the item has been indicated.
- ❒ Art has been conceptualized and sketched for the item, if applicable.
- ❒ The passage/stimulus associated with the item has been provided.

Checklist for Item Development

- ❒ The item matches content and format specifications.
- ❒ The item deals with material that is important in testing the appropriate strand.
- ❒ The item is free of gender, ethnic, racial, or other bias.
- ❒ The content of the item is grade-appropriate.
- ❒ The thinking skills demanded of the student are grade-appropriate.
- ❒ The reading level of the item strand and answer choices are suitable for the student being tested.
- ❒ All factual information has been checked and documented against reliable, up-to-date sources.
- ❒ The student can answer the question or complete the statement without looking at the answer choices.
- ❒ A student possessing the skill being tested can clearly select one and only one correct response.
- ❒ All item distractors are plausible to someone who has not mastered the skill being measured.
- ❒ The item stem presents only one question or statement.
- ❒ The item stem does not present clues to the correct response of the item.
- ❒ The item (stem and/or answer choices) does not present clues to the correct response to any other item that is in the same set of choices.
- ❒ All extraneous material has been edited from the stem.
- ❒ Answer choices are free of repetitious words or expressions that can be included in the stem.
- ❒ All answer choices are consistent with the stem both conceptually and grammatically as well as consistent with each other.
- ❒ All answer choices in the item are approximately equal in length (i.e., no one choice is much longer or shorter than another; in math, from low to high or vice-versa).
- ❒ All answer choices are mutually exclusive.
- ❒ No outliners (responses that are obviously different from the others):
  - ❒ Responses all similar in meaning.
  - ❒ Responses either all similar in length or two are long and two are short.
- ❒ Answer choices should not all begin with the same word - if this happens, include the word or words in the stem.
- ❒ Items phrased clearly and simply (check words that you suspect are too difficult a reading level against some word list).
- ❒ Check for similarity of items, repeated items, or items that give clues to other items.
- ❒ Check whether any material is copyrighted and, if so, indicate source so permission can be obtained.
- ❒ Reasonable representation of economic classes, races, ages, sexes, and handicapped in text and art:
  - ❒ Variety of above graphics.

- ❒ Non-stereotypic representation.
- ❒ Watch middle- and upper-economic level bias.

❒ Check to see that opinions are not masquerading as facts.

❒ Junk food?

❒ Is the material too dated for audience?

❒ The negative form of the stem has been used only if absolutely necessary.

❒ Key words (e.g., best, first, not, etc.) are formatted according to specifications (underlined, capitalized, italicized, left alone).

❒ The correct response for the item has been indicated.

❒ Art has been conceptualized for the item, if applicable.

❒ Position and type of art is indicated.

❒ Each piece of art is described in words and/or pictures.

❒ Descriptions of each piece of art are specific and unambiguous.

❒ Rules are clear, straight, of desired width and length. Sides drawn proportionally.

❒ Art has been checked against the corresponding item. Art or item has been revised, if necessary.

❒ Figures and tables are accurate, factual, and documented if appropriate.

❒ Males and females are represented equally in the art.

❒ Ethnic groups are represented equitably and non-stereotypically in the art.

❒ The passage/stimulus/graphic associated with the item has been indicated.

NOTE: Use your project checklist in addition to this checklist.

Sign Off

_____ _____

Name Date

Checklist for Scoring Rubrics/Scoring Guide

- ❒ Type of scoring for each scorable unit has been identified.
- ❒ A scoring rubric has been identified for each scorable unit prior to or simultaneously with item development.
- ❒ The performance criterion (outcome/strand to be assessed) has been identified for each scorable unit.
- ❒ All foreseeable correct responses have been identified.
- ❒ A scale (no. of points) has been identified for each scorable unit.
- ❒ Score points have been defined for each scorable unit (e.g., 4 = outstanding).
- ❒ Score points are clearly distinguishable from one another.
- ❒ The rubric allows full credit for answers dependent on earlier responses, even if the earlier response is incorrect.
- ❒ When more than one student behavior is required by an activity, the rubric clearly distinguishes among the behaviors and indicates how each is to be scored.
- ❒ The rubric focuses on performance (i.e., what the student did) and not on the performer (i.e., what the student understands).
- ❒ The language of the rubric is clear, consistent, and unambiguous.
- ❒ Any changes to scoring rubrics have been checked against the corresponding item.
- ❒ Scoring rubrics have been revised if any revisions occurred in the corresponding item.

Sign Off

_____ _____ _____

Name Date

Scoring Guide for Communication Arts: Reading

4 **Knowledge:** The student's response demonstrates a synthesis of relevant knowledge within and across **three** reading selections. It reveals depth and insightful connections without misconceptions about the reading selections, the scenario, and/or the scenario question.

Application: The student responds to the scenario question, stating a clear position which is effectively supported by examples from within three and across at **least two** of the reading selections.

Specifically, the student must:

- support the stated position with referenced examples from each reading selection;
- present a clear and thoughtful application of the common ideas, principles, and generalizations that connect all the reading selections.

3 **Knowledge:** The student's response demonstrates an understanding of the relevant knowledge used within and across at **least two** of the reading selections, but may NOT provide direct connections to the common ideas, generalizations or principles that tie the reading selections together. There may be minor misconceptions about the reading selections, the scenario, or the scenario question.

Application: The student responds to the scenario question, stating a clear position which is effectively supported by examples from within and across at **least two** of the reading selections.

Specifically, the student's response must:

- support the stated position with referenced examples from two reading selections
- present a clear and thoughtful application that demonstrates some understanding of the common ideas, principles, and generalizations from the reading selections to the scenario question.

2 **Knowledge:** The student response demonstrates limited or vague connections and/or insignificant references within and across the reading selections. In fact, the response may draw support from only **one** reading selection. Prior knowledge may be present, but may result in digression or provide limited support to the student's position. There may be major misconceptions about the reading selections, the scenario, and the scenario question.

Application: The student responds to the scenario question, but does not state a clear position or does not support it effectively.

Specifically, the student's response:

- provides limited examples from within and across the reading selections
- references only one reading selection
- demonstrates unclear reading application of the relationships among the common ideas, principles, or generalizations from the reading selections, the scenario, and the scenario question.

1 The student response makes reference to the reading selections, but demonstrates only a very superficial understanding of the selections and/or their relationship to the

common ideas, principles or generalizations from the reading selections, the scenario and/or the scenario question.

Condition Codes

The following condition codes will be applied to student responses that cannot be scored:

A = no reference to the scenario/answered the focus question
B = off-topic
C = illegible/written in a language other than English
D = blank/refused to respond
E = off-task (student did NOT reference any of the three reading selections in his/her written response)

Appendix B

Table 8. Michigan HSPT in Communication Arts: Reading Tryout
Raw Score Statistics by Form

Grp	Form	# Scored Items	N	Raw Score Mean	%MS[3]	SD	α	P-Value[1] 90th	P-Value[1] 10th	IT[2] 90th	IT[2] 10th	Collapsed Levels Item #	Collapsed Levels From	Collapsed Levels To
1	1	35[4]	521	24.8	60	7.5	.80	.83	.40	.49	.14	–	–	–
5		35	479	22.9	56	8.7	.85					–	–	–
1	2	35	505	23.7	60	6.5	.73	.84	.35	.41	.05	–	–	–
5		35	491	21.5	53	7.5	.80					–	–	–
1	3	35	492	26.6	65	6.9	.81	.88	.45	.49	.24	–	–	–
2		35	431	23.1	56	8.5	.88					–	–	–
2	4	35	407	22.3	54	8.2	.87	.84	.40	.52	.15	–	–	–
6		35	549	25.8	63	7.3	.84					–	–	–
2	5	36	411	18.0	44	8.2	.85	.77	.32	.45	.18	–	–	–
3		36	461	20.7	51	8.1	.84					–	–	–
3	6	36	455	21.5	52	8.3	.84	.83	.36	.50	.16	–	–	–
6		36	547	22.5	55	8.1	.84					–	–	–
3	7	36	425	19.5	48	8.2	.86	.69	.34	.48	.15	–	–	–
4		36	477	19.1	47	7.9	.85					–	–	–
4	8	36	477	19.0	46	7.1	.82	.75	.30	.44	.16	–	–	–
6		36	529	20.6	50	7.7	.84					–	–	–
4	9	36	471	23.8	58	7.5	.82	.88	.33	.47	.17	–	–	–
6		36	515	25.7	63	7.6	.83					–	–	–
4	10	36	462	22.6	55	8.2	.87	.85	.26	.49	.17	–	–	–
5		36	489	22.2	54	7.8	.87					–	–	–

1. P-values for 90th and 10th percentile when items are sorted in order of p-values.
2. Items/test correlations for 90th and 10th percentile items.
3. Mean divided by maximum score (percentage of maximum score).
4. One item in Forms 1-4 each has multiple correct answers. Therefore, these items were not included in the analysis.

Table 9. HSPT in Communication Arts: Reading Tryout
Item Means and Standard Deviations by Form

Form 1

Item	N	Mean	Std Dev
1	1015	0.90	0.30
2	1016	0.86	0.34
3	1016	0.67	0.47
4	1016	0.37	0.48
5	1016	0.70	0.46
6	1013	0.51	0.50
7	1016	0.51	0.50
8	1009	0.90	0.30
9	1015	0.65	0.48
10	1015	0.47	0.50
11	1015	0.67	0.47
12	1013	0.71	0.46
13	1014	0.69	0.46
14	1014	0.18	0.39
15	1013	0.78	0.41
16	1012	0.63	0.48
17	1011	0.74	0.44
18	1010	0.78	0.41
19	1011	0.83	0.37
20	1011	0.69	0.46
21	986	0.75	0.43
22	989	0.70	0.46
23	988	0.40	0.49
24	988	0.62	0.49
25	989	0.40	0.49
26	991	0.58	0.49
27	991	0.61	0.49
28	991	0.81	0.39
29	990	0.62	0.48
30	989	0.56	0.50
31	992	0.51	0.50
32	989	0.50	0.50
33	991	0.66	0.47
34	989	0.31	0.46
35	991	0.59	0.49
36*	934	2.88	2.35

* Extended response

Table 9 (cont). HSPT in Communication Arts: Reading Tryout
Item Means and Standard Deviations by Form

Form 2

Item	N	Mean	Std Dev
1	1010	0.88	0.33
2	1012	0.87	0.33
3	1012	0.36	0.48
4	1012	0.77	0.42
5	1011	0.79	0.41
6	1009	0.33	0.47
7	1010	0.16	0.37
8	1010	0.90	0.30
9	1008	0.68	0.47
10	1010	0.83	0.37
11	1010	0.34	0.47
12	1007	0.55	0.50
13	1010	0.37	0.48
14	1010	0.37	0.48
15	1009	0.84	0.37
16	1008	0.58	0.49
17	1009	0.68	0.47
18	1008	0.79	0.41
19	1010	0.67	0.47
20	1009	0.40	0.49
21	984	0.77	0.42
22	983	0.69	0.46
23	982	0.42	0.49
24	983	0.57	0.50
25	985	0.39	0.49
26	986	0.50	0.50
27	988	0.55	0.50
28	990	0.80	0.40
29	983	0.59	0.49
30	986	0.52	0.50
31	989	0.54	0.50
32	987	0.78	0.42
33	986	0.84	0.36
34	984	0.23	0.42
35	987	0.58	0.49
36*	916	2.70	2.31

* Extended response

Table 9 (cont). HSPT in Communication Arts: Reading Tryout
Item Means and Standard Deviations by Form

Form 3

Item	N	Mean	Std Dev
1	940	0.76	0.42
2	941	0.92	0.28
3	938	0.70	0.46
4	938	0.77	0.42
5	940	0.45	0.50
6	939	0.75	0.44
7	941	0.89	0.32
8	939	0.78	0.42
9	940	0.92	0.26
10	940	0.75	0.43
11	940	0.88	0.33
12	938	0.70	0.46
13	939	0.84	0.36
14	932	0.86	0.34
15	930	0.55	0.50
16	930	0.49	0.50
17	931	0.28	0.45
18	930	0.75	0.43
19	928	0.64	0.48
20	929	0.50	0.50
21	902	0.51	0.50
22	903	0.63	0.48
23	903	0.76	0.43
24	905	0.85	0.36
25	901	0.70	0.46
26	904	0.85	0.36
27	904	0.35	0.48
28	902	0.61	0.49
29	904	0.59	0.49
30	905	0 75	0.43
31	905	0.36	0.48
32	906	0.71	0.45
33	904	0.48	0.50
34	907	0.81	0.39
35	906	0.83	0.38
36*	847	2.17	2.02

* Extended response

Table 9 (cont). HSPT in Communication Arts: Reading Tryout
Item Means and Standard Deviations by Form

Form 4

Item	N	Mean	Std Dev
1	971	0.68	0.46
2	973	0.30	0.46
3	974	0.54	0 50
4	974	0.55	0 50
5	974	0.48	0.50
6	974	0.72	0.45
7	971	0.91	0.28
8	966	0.74	0.44
9	971	0.74	0.44
10	968	0.77	0.42
11	970	0.86	0.35
12	969	0.80	0.40
13	969	0.85	0.36
14	964	0.83	0.37
15	963	0.77	0.42
16	964	0.74	0.44
17	965	0.27	0.44
18	964	0.71	0.45
19	963	0.58	0.49
20	964	0.84	0.36
21	954	0.51	0.50
22	955	0.56	0.50
23	954	0.68	0.47
24	954	0.78	0.41
25	953	0.69	0.46
26	952	0.81	0.39
27	952	0.37	0.48
28	951	0.57	0.50
29	951	0.59	0.49
30	952	0.77	0.42
31	950	0.40	0.49
32	951	0.72	0.45
33	950	0.47	0.50
34	949	0.71	0.45
35	950	0.84	0.37
36*	892	2.09	1.85

* Extended response

Table 9 (cont). HSPT in Communication Arts: Reading Tryout
Item Means and Standard Deviations by Form

Form 5

Item	N	Mean	Std Dev
1	853	0.45	0.50
2	855	0.51	0.50
3	853	0.64	0.48
4	855	0.78	0.42
5	854	0.46	0.50
6	855	0.79	0.41
7	855	0.45	0.50
8	851	0.79	0.41
9	848	0.35	0.48
10	850	0.68	0.46
11	848	0.28	0.45
12	846	0.52	0.50
13	849	0.50	0.50
14	838	0.51	0.50
15	839	0.45	0.50
16	837	0.49	0.50
17	838	0.42	0.49
18	839	0.53	0.50
19	838	0.34	0.47
20	837	0.17	0.38
21	819	0.77	0.42
22	822	0.65	0.48
23	820	0.32	0.47
24	822	0.38	0.49
25	821	0.57	0.50
26	818	0.36	0.48
27	817	0.76	0.43
28	819	0.83	0.38
29	818	0.51	0.50
30	817	0.57	0.50
31	818	0.62	0.49
32	816	0.58	0.49
33	817	0.58	0.49
34	816	0.48	0.50
35	816	0.68	0.47
36*	746	1.70	2.06

* Extended response

Table 9 (cont). HSPT in Communication Arts: Reading Tryout
Item Means and Standard Deviations by Form

Form 6

Item	N	Mean	Std Dev
1	1012	0.48	0.50
2	1014	0.59	0.49
3	1012	0.73	0.45
4	1013	0.85	0.36
5	1012	0.82	0.38
6	1011	0.51	0.50
7	1013	0.46	0.50
8	1013	0.83	0.37
9	1009	0.41	0.49
10	1013	0.72	0.45
11	1012	0.64	0.48
12	1012	0.68	0.47
13	1013	0.47	0.50
14	1003	0.58	0.49
15	1008	0.86	0.35
16	1005	0.84	0.37
17	1006	0.66	0.47
18	1006	0.62	0.48
19	1003	0.30	0.46
20	1003	0.24	0.42
21	997	0.78	0.42
22	994	0.62	0.48
23	992	0.43	0.50
24	992	0.47	0.50
25	989	0.59	0.49
26	990	0.41	0.49
27	992	0.48	0.50
28	991	0.33	0.47
29	991	0.54	0.50
30	990	0.58	0.49
31	989	0.65	0.48
32	989	0.59	0.49
33	990	0.59	0.49
34	990	0.46	0.50
35	990	0.69	0.46
36*	934	2.14	2.17

* Extended response

Table 9 (cont). HSPT in Communication Arts: Reading Tryout
Item Means and Standard Deviations by Form

Form 7

Item	N	Hean	Std Dev
1	890	0.40	0.49
2	890	0.48	0.50
3	890	0.50	0.50
4	890	0.74	0.44
5	890	0.76	0.43
6	889	0.61	0.49
7	890	0.51	0.50
8	887	0.25	0.44
9	890	0.34	0.47
10	890	0.57	0.50
11	881	0.51	0.50
12	879	0.22	0.41
13	881	0.64	0.48
14	882	0.65	0.48
15	881	0.54	0.50
16	872	0.69	0.46
17	878	0.34	0.47
18	877	0.53	0.50
19	877	0.46	0.50
20	877	0.43	0.50
21	855	0.66	0.47
22	857	0.74	0.44
23	855	0.63	0.48
24	855	0.52	0.50
25	854	0.37	0.48
26	853	0.38	0.49
27	854	0.39	0.49
28	855	0.59	0.49
29	855	0.42	0.49
30	855	0.63	0.48
31	855	0.73	0.45
32	854	0.59	0.49
33	854	0.59	0.49
34	853	0.57	0.50
35	853	0.61	0.49
36*	791	1.56	1.78

* Extended response

Table 9 (cont). HSPT in Communication Arts: Reading Tryout
Item Means and Standard Deviations by Form

Form 8

Item	N	Mean	Std Dev
1	1012	0.41	0.49
2	1016	0.48	0.50
3	1016	0.59	0.49
4	1016	0.76	0.42
5	1016	0.75	0.43
6	1019	0.63	0.48
7	1014	0.58	0.49
8	1017	0.29	0.45
9	1017	0.34	0.47
10	1016	0.55	0.50
11	1014	0.49	0.50
12	1015	0.23	0.42
13	1016	0.82	0.38
14	1014	0.69	0.46
15	1014	0.32	0.47
16	1001	0.74	0.44
17	1004	0.39	0.49
18	1004	0.20	0.40
19	1001	0.47	0.50
20	998	0.43	0.49
21	996	0.85	0.36
22	996	0.50	0.50
23	996	0.66	0.48
24	996	0.47	0.50
25	990	0.36	0.48
26	989	0.38	0.49
27	989	0.43	0.50
28	990	0.56	0.50
29	990	0.47	0.50
30	990	0.62	0.49
31	987	0.60	0.49
32	990	0.77	0.42
33	988	0.71	0.46
34	989	0.56	0.50
35	987	0.58	0.49
36*	933	1.79	1.73

* Extended response

Table 9 (cont). HSPT in Communication Arts: Reading Tryout
Item Means and Standard Deviations by Form

Form 9

Item	N	Mean	Std Dev
1	1002	0.70	0.46
2	1001	0.67	0.47
3	1001	0.55	0.50
4	1003	0.94	0.24
5	1002	0.86	0.35
6	1003	0.58	0.49
7	996	0.25	0.43
8	996	0.71	0.45
9	998	0.33	0.47
10	998	0.50	0.50
11	997	0.72	0.45
12	999	0.89	0.31
13	999	0.66	0.47
14	999	0.86	0.35
15	993	0.72	0.45
16	992	0.73	0.44
17	991	0.73	0.44
18	994	0.48	0.50
19	990	0.81	0.40
20	992	0.77	0.42
21	977	0.59	0.49
22	978	0.91	0.29
23	974	0.72	0.45
24	977	0.88	0.33
25	978	0.44	0.50
26	978	0.48	0.50
27	973	0.16	0.37
28	977	0.62	0.49
29	975	0.68	0.47
30	977	0.82	0.38
31	975	0.70	0.46
32	975	0.77	0.42
33	975	0.88	0.32
34	975	0.39	0.49
35	974	0.28	0.45
36*	937	2.63	2.10

* Extended response

Table 9 (cont). HSPT in Communication Arts: Reading Tryout
Item Means and Standard Deviations by Form

Form 10

Item	N	Mean	Std Dev
1	930	0.67	0.47
2	932	0.89	0.32
3	930	0.53	0.50
4	931	0.91	0.28
5	932	0.84	0.37
6	931	0.77	0.42
7	930	0.25	0.44
8	927	0.68	0.47
9	930	0.32	0.47
10	931	0.46	0.50
11	929	0.58	0.49
12	930	0.84	0.37
13	930	0.59	0.49
14	930	0.82	0.38
15	926	0.65	0.48
16	926	0.64	0.48
17	924	0.63	0.48
18	924	0.44	0.50
19	924	0.76	0.43
20	923	0.68	0.47
21	901	0.56	0.50
22	905	0.89	0.31
23	900	0.67	0.47
24	905	0.87	0.33
25	902	0.44	0.50
26	900	0.43	0.50
27	902	0.16	0.37
28	898	0.54	0.50
29	900	0.64	0.48
30	902	0.77	0.42
31	901	0.66	0.48
32	901	0.73	0.44
33	899	0.85	0.36
34	899	0.38	0.49
35	902	0.25	0.43
36*	853	1.57	1.64

* Extended response

Table 10. HSPT in Communication Arts: Reading Tryout
Item P-Values of Quintiles by Form

Form 1

Item in Test	P-value Quintile 1*	P-value Quintile 2	P-value Quintile 3	P-value Quintile 4	P-value Quintile 5
1	0.99	0.99	0.95	0.90	0.70
2	0.98	0.97	0.93	0.86	0.61
3	0.94	0.81	0.70	0.56	0.40
4	0.70	0.47	0.31	0.20	0.21
5	0.85	0.80	0.76	0.68	0.43
6	0.58	0.59	0.49	0.52	0.39
7	0.76	0.58	0.48	0.42	0.33
8	0.99	0.98	0.97	0.93	0.62
9	0.77	0.73	0.70	0.65	0.43
10	0.75	0.53	0.48	0.40	0.25
11	0.97	0.87	0.74	0.54	0.28
12	0.97	0.89	0.79	0.59	0.33
13	0.93	0.85	0.76	0.54	0.37
14	0.17	0.19	0.19	0.19	0.17
15	0.95	0.92	0.87	0.77	0.43
16	0.88	0.74	0.63	0.57	0.37
17	0.97	0.92	0.79	0.70	0.36
18	0.99	0.92	0.92	0.78	0.31
19	0.98	0.97	0.93	0.87	0.44
20	0.83	0.77	0.77	0.71	0.37
21	0.94	0.81	0.81	0.67	0.47
22	0.90	0.81	0.70	0.60	0.44
23	0.52	0.46	0.39	0.33	0.26
24	0.97	0.80	0.68	0.39	0.24
25	0.54	0.43	0.41	0.37	0.21
26	0.91	0.77	0.62	0.40	0.18
27	0.94	0.72	0.63	0.45	0.26
28	0.99	0.96	0.93	0.73	0.38
29	0.90	0.75	0.62	0.55	0.27
30	0.91	0.74	0.58	0.33	0.19
31	0.85	0.64	0.52	0.35	0.20
32	0.75	0.64	0.47	0.37	0.23
33	0.92	0.85	0.73	0.53	0.23
34	0.55	0.38	0.21	0.19	0.21
35	0.74	0.68	0.61	0.55	0.31
Extended-response item #36	5.38	3.94	2.55	1.44	0.50

* P-values Quintile 1 - students' scores between the 81st-100th percentiles;
P-values Quintile 2 - students' scores between the 61st-80th percentiles and so on.

Table 10 (cont). HSPT in Communication Arts: Reading Tryout
Item P-Values of Quintiles by Form

Form 2

Item in Test	P-value Quintile 1	P-value Quintile 2	P-value Quintile 3	P-value Quintile 4	P-value Quintile 5
1	0.99	0.94	0.95	0.84	0.69
2	0.97	0.96	0.95	0.89	0.61
3	0.45	0.39	0.32	0.33	0.29
4	0.95	0.87	0.81	0.70	0.52
5	0.95	0.91	0.83	0.77	0.50
6	0.47	0.33	0.32	0.27	0.24
7	0.22	0.17	0.14	0.16	0.14
8	0.97	0.98	0.96	0.92	0.68
9	0.93	0.84	0.73	0.54	0.36
10	0.96	0.92	0.88	0.85	0.57
11	0.49	0.41	0.33	0.26	0.19
12	0.87	0.69	0.56	0.38	0.25
13	0.64	0.40	0.27	0.29	0.24
14	0.42	0.39	0.42	0.40	0.23
15	1.00	0.98	0.92	0.81	0.48
16	0.82	0.65	0.60	0.49	0.36
17	0.83	0.78	0.61	0.67	0.45
18	0.96	0.94	0.88	0.80	0.37
19	0.88	0.83	0.65	0.62	0.35
20	0.63	0.39	0.41	0.36	0.25
21	0.94	0.85	0.76	0.73	0.46
22	0.84	0.74	0.66	0.62	0.47
23	0.55	0.44	0.39	0.40	0.28
24	0.84	0.74	0.48	0.39	0.28
25	0.51	0.42	0.35	0.38	0.23
26	0.81	0.66	0.47	0.31	0.16
27	0.86	0.69	0.51	0.34	0.28
28	0.98	0.95	0.90	0.76	0.37
29	0.82	0.66	0.63	0.53	0.25
30	0.82	0.68	0.49	0.35	0.18
31	0.83	0.65	0.57	0.37	0.24
32	0.95	0.90	0.90	0.73	0.35
33	0.98	0.96	0.94	0.82	0.43
34	0.30	0.23	0.22	0.19	0.17
35	0.71	0.61	0.66	0.57	0.30
Extended-response item #36	5.19	3.72	2.17	1.31	0.48

Table 10 (cont). HSPT in Communication Arts: Reading Tryout
Item P-Values of Quintiles by Form

Form 3

Item in Test	P-value Quintile 1	P-value Quintile 2	P-value Quintile 3	P-value Quintile 4	P-value Quintile 5
1	0.94	0.85	0.79	0.67	0.56
2	0.98	0.99	0.95	0.94	0.72
3	0.91	0.79	0.74	0.62	0.43
4	0.98	0.91	0.84	0.69	0.39
5	0.50	0.37	0.44	0.46	0.45
6	0.89	0.84	0.80	0.70	0.48
7	1.00	0.99	0.99	0.86	0.60
8	0.96	0.90	0.80	0.72	0.49
9	1.00	1.00	1.00	0.96	0.66
10	0.94	0.89	0.78	0.69	0.44
11	1.00	0.96	0.93	0.90	0.60
12	0.88	0.84	0.76	0.61	0.37
13	0.96	0.93	0.94	0.84	0.53
14	1.00	0.98	0.97	0.84	0.48
15	0.82	0.65	0.54	0.40	0.28
16	0.76	0.58	0.44	0.30	0.30
17	0.37	0.28	0.24	0.26	0.22
18	0.96	0.88	0.78	0.69	0.38
19	0.89	0.84	0.66	0.49	0.27
20	0.83	0.61	0.43	0.32	0.26
21	0.70	0.56	0.54	0.39	0.26
22	0.72	0.69	0.68	0.64	0.29
23	0 92	0.88	0.85	0.67	0.29
24	0 99	0.99	0.94	0.78	0.37
25	0.89	0.84	0.76	0.64	0.21
26	0.98	0.97	0.95	0.80	0.36
27	0.70	0.37	0.23	0.21	0.12
28	0.89	0.77	0.62	0.40	0.24
29	0.83	0.66	0.58	0.42	0.30
30	0.97	0.87	0.82	0.65	0.30
31	0.68	0.41	0.29	0.20	0.15
32	0.93	0.89	0.77	0.54	0.28
33	0.72	0.59	0.40	0.40	0.21
34	0.98	0.97	0.90	0.73	0.31
35	0.99	0.96	0.92	0.81	0.30
Extended-response item #36	4.60	2.92	1.69	0.94	0.28

Table 10 (cont). HSPT in Communication Arts: Reading Tryout
Item P-Values of Quintiles by Form

Form 4

Item in Test	P-value Quintile 1	P-value Quintile 2	P-value Quintile 3	P-value Quintile 4	P-value Quintile 5
1	0.90	0.80	0.73	0.55	0.41
2	0.45	0.29	0.25	0.22	0.30
3	0.82	0.67	0.50	0.43	0.27
4	0.69	0.59	0.50	0.55	0.39
5	0.56	0.44	0.47	0.49	0.43
6	0.89	0.78	0.75	0.65	0.50
7	1.00	0.99	0.99	0.96	0.59
8	0.95	0.87	0.77	0.64	0.40
9	0.93	0.91	0.75	0.65	0.41
10	0.96	0.91	0.81	0.75	0.37
11	0.99	0.96	0.91	0.86	0.51
12	0.98	0.95	0.87	0.72	0.43
13	0.98	0.98	0.93	0.83	0.45
14	0.98	0.98	0.94	0.79	0.40
15	0.98	0.94	0.83	0.71	0.32
16	0.93	0.87	0.77	0.67	0.40
17	0.45	0.25	0.28	0.25	0.14
18	0.94	0.88	0.75	0.56	0.32
19	0.82	0.69	0.61	0.46	0.26
20	0.99	0.97	0.91	0.81	0.46
21	0.72	0.62	0.48	0.43	0.23
22	0.70	0.61	0.61	0.50	0.32
23	0.89	0.82	0.70	0.59	0.27
24	0.97	0.96	0.84	0.69	0.32
25	0.93	0.83	0.77	0.58	0.25
26	0.98	0.96	0.93	0.77	0.30
27	0.71	0.48	0.27	0.20	0.12
28	0.87	0.67	0.53	0.43	0.27
29	0.85	0.68	0.62	0.43	0.26
30	0.98	0.91	0.83	0.68	0.31
31	0.74	0.48	0.30	0.22	0.18
32	0.94	0.88	0.78	0.57	0.28
33	0.74	0.56	0.40	0.30	0.25
34	0.95	0.90	0.78	0.58	0.19
35	0.99	0.98	0.96	0.77	0.35
Extended-response item #36	4.23	2.77	1.71	1.11	0.32

Table 10 (cont). HSPT in Communication Arts: Reading Tryout
Item P-Values of Quintiles by Form

Form 5

Item in Test	P-value Quintile 1	P-value Quintile 2	P-value Quintile 3	P-value Quintile 4	P-value Quintile 5
1	0.76	0.50	0.43	0.33	0.21
2	0.77	0.68	0.41	0.36	0.30
3	0.95	0.80	0.67	0.44	0.30
4	0.98	0.94	0.85	0.64	0.44
5	0.79	0.48	0.36	0.36	0.29
6	0.93	0.89	0.85	0.73	0.49
7	0.74	0.60	0.43	0.26	0.21
8	0.96	0.96	0.81	0.68	0.48
9	0.55	0.38	0.39	0.22	0.20
10	0.90	0.80	0.73	0.57	0.37
11	0.33	0.36	0.17	0.27	0.24
12	0.77	0.61	0.46	0.41	0.28
13	0.78	0.60	0.42	0.33	0.32
14	0.81	0.65	0.53	0.32	0.16
15	0.69	0.53	0.42	0.36	0.19
16	0.75	0.54	0.46	0.39	0.24
17	0.62	0.49	0.38	0.33	0.21
18	0.87	0.66	0.47	0.32	0.21
19	0.63	0.35	0.26	0.21	0.18
20	0.35	0.13	0.14	0.11	0.09
21	0.97	0.92	0.84	0.62	0.30
22	0.88	0.85	0.69	0.50	0.16
23	0.52	0.33	0.30	0.28	0.08
24	0.61	0.46	0.35	0.26	0.15
25	0.86	0.69	0.56	0.40	0.21
26	0.57	0.27	0.30	0.30	0.21
27	0.94	0.93	0.82	0.64	0.25
28	0.99	0.99	0.96	0.70	0.28
29	0.83	0.73	0.48	0.26	0.10
30	0.79	0.71	0.54	0.45	0.18
31	0.85	0.78	0.65	0.43	0.24
32	0.84	0.68	0.59	0.43	0.20
33	0.90	0.85	0.55	0.34	0.08
34	0.61	0.53	0.55	0.38	0.18
35	0.94	0.83	0.78	0.46	0.21
Extended-response item #36	4.15	2.24	1.01	0.33	0.15

Table 10 (cont). HSPT in Communication Arts: Reading Tryout
Item P-Values of Quintiles by Form

Form 6

Item in Test	P-value Quintile 1	P-value Quintile 2	P-value Quintile 3	P-value Quintile 4	P-value Quintile 5
1	0.85	0.60	0.43	0.33	0.22
2	0.90	0.70	0.63	0.42	0.31
3	0.97	0.89	0.75	0.62	0.40
4	0.97	0.96	0.91	0.84	0.55
5	0.97	0.96	0.90	0.76	0.52
6	0.71	0.62	0.46	0.44	0.30
7	0.82	0.50	0.44	0.28	0.30
8	0.96	0.94	0.89	0.80	0.58
9	0.62	0.49	0.36	0.35	0.22
10	0.87	0.84	0.80	0.72	0.39
11	0.86	0.82	0.71	0.57	0.23
12	0.90	0.83	0.71	0.59	0.39
13	0.80	0.57	0.43	0.32	0.24
14	0.87	0.73	0.56	0.43	0.28
15	0.99	0.98	0.92	0.86	0.51
16	1.00	0.98	0.94	0.82	0.41
17	0.85	0.75	0.66	0.59	0.41
18	0.90	0.73	0.62	0.56	0.30
19	0.49	0.34	0.25	0.23	0.19
20	0.48	0.26	0.16	0.13	0.16
21	0.99	0.95	0.89	0.63	0.36
22	0.88	0.74	0.63	0.56	0.28
23	0.80	0.50	0.31	0.25	0.28
24	0.83	0.59	0.44	0.28	0.17
25	0.86	0.69	0.58	0.48	0.30
26	0.78	0.46	0.30	0.26	0.20
27	0.63	0.54	0.44	0.44	0.30
28	0.44	0.42	0.30	0.26	0.19
29	0.85	0.74	0.53	0.37	0.14
30	0.79	0.69	0.66	0.45	0.25
31	0.85	0.78	0.74	0.56	0.27
32	0.89	0.72	0.63	0.43	0.23
33	0.96	0.78	0.68	0.34	0.15
34	0.65	0.52	0.50	0.39	0.21
35	0.95	0.90	0.80	0.55	0.21
Extended-response item #36	4.64	2.93	1.83	0.80	0.20

Table 10 (cont). HSPT in Communication Arts: Reading Tryout
Item P-Values of Quintiles by Form

Form 7

Item in Test	P-value Quintile 1	P-value Quintile 2	P-value Quintile 3	P-value Quintile 4	P-value Quintile 5
1	0.60	0.50	0.38	0.32	0.17
2	0.76	0.63	0.47	0.34	0.19
3	0.88	0.71	0.42	0.29	0.17
4	0.97	0.90	0.81	0.65	0.35
5	0.99	0.94	0.84	0.66	0.35
6	0.96	0.75	0.57	0.46	0.31
7	0.79	0.66	0.47	0.32	0.33
8	0.40	0.23	0.23	0.19	0.22
9	0.63	0.44	0.31	0.16	0.16
10	0.78	0.76	0.56	0.49	0.25
11	0.75	0.56	0.50	0.41	0.30
12	0.41	0.20	0.16	0.13	0.18
13	0.95	0.80	0.66	0.52	0.23
14	0.90	0.79	0.70	0.50	0.34
15	0.81	0.64	0.55	0.39	0.25
16	0.96	0.89	0.76	0.49	0.29
17	0.64	0.36	0.32	0.19	0.18
18	0.87	0.72	0.50	0.30	0.19
19	0.65	0.56	0.50	0.33	0.20
20	0.70	0.49	0.40	0.27	0.25
21	0.79	0.74	0.72	0.63	0.27
22	0.99	0.89	0.86	0.56	0.23
23	0.90	0.75	0.60	0.50	0.27
24	0.75	0.48	0.46	0.49	0.28
25	0.47	0.42	0.34	0.33	0.21
26	0.54	0.39	0.34	0.32	0.21
27	0.61	0.47	0.38	0.27	0.11
28	0.90	0.74	0.50	0.45	0.20
29	0.63	0.53	0.35	0.31	0.18
30	0.93	0.75	0.62	0.50	0.21
31	0.95	0.89	0.78	0.59	0.25
32	0.90	0.82	0.57	0.35	0.20
33	0.86	0.81	0.63	0.39	0.13
34	0.80	0.64	0.59	0.45	0.26
35	0.86	0.73	0.58	0.48	0.24
Extended-response item #36	3.72	1.87	1.22	0.49	0.16

Table 10 (cont). HSPT in Communication Arts: Reading.Tryout
Item P-Values of Quintiles by Form

Form 8

Item in Test	P-value Quintile 1	P-value Quintile 2	P-value Quintile 3	P-value Quintile 4	P-value Quintile 5
1	0.65	0.48	0.43	0.24	0.23
2	0.75	0.65	0.50	0.32	0.19
3	0.84	0.70	0.61	0.48	0.30
4	0.96	0.93	0.83	0.63	0.46
5	0.98	0.93	0.84	0.61	0.39
6	0.94	0.80	0.66	0.44	0.32
7	0.78	0.63	0.58	0.57	0.34
8	0.45	0.33	0.22	0.25	0.20
9	0.57	0.45	0.31	0.19	0.17
10	0.80	0.72	0.53	0.41	0.29
11	0.82	0.60	0.47	0.34	0.20
12	0.43	0.20	0.21	0.15	0.17
13	0.99	0.93	0.89	0.79	0.49
14	0.92	0.80	0.67	0.64	0.40
15	0.59	0.37	0.26	0.20	0.20
16	0.97	0.91	0.76	0.62	0.38
17	0.62	0.46	0.38	0.31	0.17
18	0.28	0.15	0.18	0.17	0.19
19	0.72	0.53	0.43	0.35	0.29
20	0.77	0.52	0.38	0.25	0.19
21	0.97	0.96	0.92	0.83	0.47
22	0.66	0.63	0.57	0.37	0.19
23	0.95	0.89	0.62	0.49	0.26
24	0.80	0.50	0.34	0.38	0.30
25	0.57	0.39	0.32	0.32	0.16
26	0.60	0.45	0.34	0.30	0.19
27	0.68	0.50	0.43	0.34	0.16
28	0.88	0.76	0.49	0.37	0.22
29	0.73	0.53	0.44	0.34	0.22
30	0.96	0.83	0.62	0.36	0.24
31	0.95	0.79	0.60	0.40	0.17
32	0.98	0.92	0.86	0.66	0.34
33	0.90	0.87	0.81	0.58	0.26
34	0.80	0.64	0.52	0.48	0.28
35	0.85	0.68	0.57	0.41	0.33
Extended-response item #36	3.57	2.28	1.64	0.94	0.34

Table 10 (cont). HSPT in Communication Arts: Reading Tryout
Item P-Values of Quintiles by Form

Form 9

Item in Test	P-value Quintile 1	P-value Quintile 2	P-value Quintile 3	P-value Quintile 4	P-value Quintile 5
1	0.90	0.71	0.70	0.60	0.57
2	0.89	0.76	0.72	0.57	0.35
3	0.81	0.61	0.52	0.49	0.30
4	1.00	1.00	0.98	0.93	0.77
5	0.98	0.95	0.91	0.85	0.60
6	0.74	0.65	0.58	0.51	0.39
7	0.39	0.20	0.23	0.24	0.19
8	0.92	0.80	0.72	0.61	0.47
9	0.68	0.33	0.27	0.15	0.16
10	0.74	0.64	0.43	0.34	0.29
11	0.94	0.82	0.68	0.60	0.48
12	0.97	0.99	0.93	0.87	0.67
13	0.85	0.74	0.70	0.54	0.41
14	0.99	0.96	0.91	0.89	0.53
15	0.94	0.90	0.77	0.62	0.30
16	0.99	0.92	0.80	0.57	0.30
17	0.99	0.87	0.76	0.57	0.36
18	0.84	0.57	0.40	0.28	0.25
19	1.00	0.96	0.87	0.69	0.42
20	0.98	0.89	0.83	0.66	0.40
21	0.84	0.62	0.59	0.51	0.30
22	1.00	1.00	0.96	0.90	0.56
23	0.94	0.84	0.71	0.66	0.34
24	0.99	0.97	0.93	0.87	0.51
25	0.70	0.53	0.35	0.29	0.21
26	0.75	0.54	0.41	0.29	0.27
27	0.12	0.14	0.12	0.20	0.23
28	0.96	0.83	0.58	0.34	0.21
29	0.96	0.83	0.76	0.49	0.22
30	0.99	0.99	0.94	0.71	0.34
31	0.92	0.86	0.70	0.62	0.30
32	0.96	0.86	0.79	0.74	0.36
33	1.00	0.99	0.96	0.88	0.47
34	0.65	0.44	0.30	0.27	0.18
35	0.37	0.32	0.25	0.28	0.16
Extended-response item #36	4.66	3.39	2.23	1.58	0.69

Table 10 (cont). HSPT in Communication Arts: Reading Tryout
Item P-Values of Quintiles by Form

Form 10

Item in Test	P-value Quintile 1	P-value Quintile 2	P-value Quintile 3	P-value Quintile 4	P-value Quintile 5
1	0.95	0.85	0.74	0.44	0.34
2	1.00	0.99	0.92	0.91	0.59
3	0.77	0.64	0.52	0.45	0.27
4	0.99	0.98	0.97	0.96	0.66
5	1.00	0.96	0.89	0.82	0.52
6	0.97	0.90	0.84	0.70	0.42
7	0.38	0.22	0.22	0.23	0.23
8	0.91	0.78	0.73	0.57	0.39
9	0.60	0.40	0.24	0.16	0.21
10	0.70	0.55	0.49	0.33	0.24
11	0.75	0.66	0.59	0.54	0.38
12	0.98	0.97	0.95	0.75	0.53
13	0.93	0.80	0.52	0.42	0.27
14	0.97	0.96	0.90	0.77	0.47
15	0.98	0.84	0.74	0.45	0.23
16	0.97	0.93	0.70	0.38	0.22
17	0.87	0.75	0.69	0.48	0.32
18	0.83	0.56	0.36	0.24	0.22
19	0.98	0.97	0.84	0.62	0.33
20	0.94	0.88	0.71	0.54	0.30
21	0.81	0.71	0.61	0.43	0.15
22	0.99	0.99	0.99	0.88	0.46
23	0.90	0.82	0.68	0.55	0.29
24	0.99	0.97	0.94	0.84	0.49
25	0.63	0.48	0.44	0.35	0.21
26	0.77	0.46	0.43	0.27	0.17
27	0.08	0.14	0.13	0.23	0.20
28	0.93	0.74	0.51	0.25	0.19
29	0.94	0.84	0.71	0.41	0.17
30	1.00	0.97	0.86	0.62	0.25
31	0.87	0.73	0.68	0.57	0.32
32	0.95	0.94	0.75	0.58	0.31
33	0.99	0.97	0.92	0.77	0.40
34	0.62	0.45	0.30	0.25	0.23
35	0.33	0.26	0.26	0.23	0.12
Extended-response item #36	3.29	1.64	1.23	0.82	0.35

Table 11. Michigan HSPT in Communication Arts: Reading Tryout
Summary of Fit Results - 1PL/PPC

Grp	Form	N	# of Scored Items	# of Misfit Items				Two Largest Zs		Unest. Items	
				Z≥10	10>Z≥5	5>Z≥3	3>Z≥2			Number	Item #
1	1	519	35	1	3	5	4	12.5	6.4	0	
5		470	35	1	7	3	6	10.0	8.4	0	
1	2	503	35	0	9	3	3	8.0	7.9	0	
5		483	35	2	7	5	2	13.6	11.6	0	
1	3	492	35	2	1	3	5	30.3	10.7	0	
2		424	35	1	2	4	6	32.3	9.1	0	
2	4	403	35	4	1	3	7	40.5	25.4	0	
6		547	35	3	5	5	4	27.5	21.2	0	
2	5	401	36	0	10	2	2	9.5	8.6	0	
3		461	36	1	2	6	5	20.5	7.9	0	
3	6	450	36	0	5	5	6	8.5	5.7	0	
6		540	36	0	8	3	5	7.1	7.0	0	
3	7	419	36	0	2	7	6	9.5	5.9	0	
4		472	36	1	4	6	5	17.8	7.0	0	
4	8	473	36	0	3	8	5	8.6	5.2	0	
6		525	36	2	5	7	3	24.2	11.5	0	
4	9	468	36	2	9	5	0	38.5	27.1	0	
6		510	36	1	7	7	2	59.6	9.5	0	
4	10	454	36	2	3	9	2	25.8	12.2	0	
5		479	36	3	5	2	6	45.6	24.2	0	

Table 12. Michigan HSPT in Communication Arts: Reading Tryout
Summary of Fit Results - 3PL/2PPC

Grp	Form	N	# of Scored Items	# of Misfit Items				Two largest Zs		Unest. Items	
				Z≥10	10≥Z≥5	5≥Z≥3	3≥Z≥2			Number	Item #
1	1	519	35	0	0	4	3	4.7	4.4	0	
5		470	35	0	1	3	3	5.0	4.3	0	
1	2	503	35	1	3	2	2	10.3	5.9	1	7
5		483	35	1	0	1	4	16.4	4.0	1	7
1	3	492	35	0	1	1	3	5.7	3.8	1	5
2		424	35	0	1	2	4	6.5	3.5	1	5
2	4	403	35	0	0	3	2	4.8	3.2	1	
6		547	35	0	0	3	0	4.5	4.3	1	
2	5	401	36	0	1	1	4	7.7	4.2	0	
3		457	36	1	1	0	2	41.3	6.0	0	
3	6	450	36	0	0	2	2	3.5	3.1	0	
6		540	36	0	0	1	4	3.9	2.7	0	
3	7	419	36	0	0	2	3	3.9	3.6	0	
4		472	36	0	0	1	1	4.5	2.1	1	8*
4	8	473	36	1	0	0	0	25.2	---	0	
6		525	36	0	2	1	2	6.6	5.3	1	18
4	9	468	36	0	0	4	0	4.3	4.1	1	27
6		510	36	0	0	3	1	4.2	3.4	2	27,35
4	10	454	36	0	1	2	2	5.1	4.4	2	27, 35
5		479	36	1	0	0	2	12.5	2.1	2	7, 27

* Item/test correlation > .08.

Table 13. Michigan HSPT in Communication Arts: Reading Tryout
Items Flagged for Deletion Under the Fit Criteria - 1PL/PPC and 3PL/2PPC

	1PL/PPC		3PL/2PPC		
Form	# Misfit Items[1]	Item Number	# Misfit Items[1]	Item Number	NC[2]
1	3	6, 23, 28	1	9	0
2	7	3[$],6[$],7[$],14[$],15,28, 33	2	3[$],14[$]	1
3	3	5[$],16, 22	0		1
4	6	2[$],4,5[$],25,34,35	0		1
5	2	11[$],28	0		0
6	6	15,16,21,27,33,35	0		0
7	5	8,12,16,22,25	0		1
8	3	8,18[$],21	0		1
9	9	1,7[$],15,22,27[$],28,30,33,35[$]	0		2
10	7	7[$],9,22,27[$],28,30,35[$]	0		2

1. Note that each item has two Zs, one from one sample and the other from a second sample. A "misfit" item is defined as follows:
(1) both Zs≥ 4.0, or
(2) (one Z≥ 4.0), and (4.0> the other Z≥ 3.0), and a plot of expected and observed curves fails to demonstrate reasonable fit.

   Of the 51 items that were not fitted by the one-parameter model, 15 items fell into the latter category, (2). Of the three items not fitted by the 3PL/2PPC model, one fell in this category.

2. Maximum number of non-convergent items in a given form taken by two samples.

$. Item/test correlation < .08 signifying low discrimination.

Table 14. Michigan HSPT in Communication Arts: Reading Tryout
Mean and Standard Deviations of Item Discrimination - 3PL/2PPC

Form	Group	All Items			Multiple-Choice Only			Extended-Response Only		
		# Items	Mean	S.D.	# Items	Mean	S.D.	# Items	Mean	S.D.
1	1	35	1.23	0.59	34	1.26	0.58	1	0.35	
1	5	35	1.54	0.69	34	1.57	0.67	1	0.38	
2	1	34	1.14	0.65	33	1.17	0.64	1	0.28	
2	5	34	1.31	0.73	33	1.34	0.72	1	0.35	
3	1	34	1.45	0.63	33	1.49	0.61	1	0.36	
3	2	34	1.69	0.66	33	1.72	0.65	1	0.64	
4	2	33	1.57	0.62	32	1.61	0.60	1	0.52	
4	6	33	1.45	0.67	32	1.48	0.66	1	0.52	
5	2	36	1.53	0.69	35	1.56	0.68	1	0.60	
5	3	36	1.43	0.62	35	1.46	0.62	1	0.57	
6	3	36	1.47	0.64	35	1.49	0.62	1	0.51	
6	6	36	1.52	0.69	35	1.55	0.68	1	0.37	
7	3	35	1.50	0.58	34	1.53	0.57	1	0.61	
7	4	35	1.40	0.58	34	1.43	0.56	1	0.41	
8	4	35	1.44	0.64	34	1.47	0.63	1	0.47	
8	6	35	1.46	0.56	34	1.49	0.54	1	0.49	
9	4	34	1.56	0.81	33	1.59	0.80	1	0.29	
9	6	34	1.55	0.72	33	1.59	0.70	1	0.29	
10	4	33	1.68	0.68	32	1.72	0.66	1	0.44	
10	5	33	1.64	0.68	32	1.68	0.65	1	0.39	
Total All Groups		690	1.48	0.67	670	1.51	0.66	20	0.44	0.11

Table 20. Michigan HSPT in Communication Arts: Reading Tryout, Mantel-Haenszel Statistics
Chi-square & Standardized Mean Difference (SMD)
Reference: White or Male/Focus: Black or Female

Form 1

Item	ethnic chi sq.	ethnic SMD	gender chi sq.	gender SMD
1	0.99	0.07	1.52	0.02
2	1.23	-0.04	2.72	-0.03
3	0.56	0.05	0.00	0.00
4	0.21	0.03	1.00	-0.04
5	0.31	0.03	1.67	0.03
6	0.25	-0.04	1.64	0.05
7	0.87	-0.08	0.32	-0.02
8	0.73	-0.06	0.02	0.00
9	0.06	0.02	0.50	-0.03
10	0.84	-0.06	1.49	-0.05
11	5.50	-0.14	0.18	-0.01
12	0.92	-0.06	0.21	0.01
13	1.67	0.06	0.00	0.00
14	0.27	-0.04	1.58	0.03
15	2.17	0.09	0.24	0.01
16	0.16	-0.02	0.01	0.01
17	1.46	-0.08	0.56	0.02
18	0.29	0.03	1.97	0.03
19	0.03	-0.02	8.24	0.05
20	0.01	0.01	1.04	0.05
21	0.49	0.05	0.17	-0.01
22	0.59	0.07	0.88	0.03
23	0.78	0.08	0.00	-0.02
24	2.21	0.09	1.68	-0.04
25	0.01	0.01	3.23	0.05
26	4.20	0.13	0.15	0.01
27	0.06	0.02	0.02	0.00
28	0.08	-0.02	1.42	0.03
29	0.04	0.02	17.24	-0.12
30	0.00	-0.02	6.63	-0.07
31	0.01	0.01	1.95	-0.04
32	0.01	0.00	3.32	-0.05
33	0.93	0.07	0.04	0.01
34	0.09	0.02	2.45	-0.04
35	0.04	0.01	0.39	0.01
36*	5.63	-0.32	15.71	0.23

Form 2

Item	ethnic chi sq.	ethnic SMD	gender chi sq.	gender SMD
1	1.02	0.05	0.75	-0.02
2	0.45	-0.03	0.19	-0.01
3	0.00	-0.00	1.25	0.03
4	0.00	0.01	5.84	0.07
5	0.08	0.02	0.96	0.02
6	1.03	0.07	0.16	0.02
7	1.16	0.05	0.12	-0.00
8	2.08	0.06	0.64	-0.01
9	1.44	0.07	4.08	-0.06
10	0.02	0.00	6.11	-0.05
11	0.23	0.03	0.29	0.01
12	2.11	-0.09	8.90	-0.09
13	0.87	-0.06	0.31	0.03
14	0.29	-0.04	1.11	-0.02
15	0.00	0.02	1.03	0.02
16	3.23	0.11	0.74	0.03
17	0.32	0.04	1.45	0.04
18	0.03	-0.02	1.66	0.02
19	0.43	-0.04	0.11	-0.01
20	0.37	-0.04	1.86	-0.03
21	0.01	-0.01	3.59	-0.05
22	1.16	-0.07	13.48	0.10
23	2.10	-0.09	0.00	0.01
24	0.01	-0.01	2.36	-0.04
25	0.01	-0.01	0.00	0.00
26	2.10	-0.08	0.00	0.00
27	0.03	0.01	5.06	-0.07
28	3.14	0.08	0.01	0.00
29	1.28	0.08	1.37	-0.05
30	0.00	-0.01	1.33	-0.10
31	1.07	0.07	0.18	-0.02
32	0.00	-0.01	0.10	-0.01
33	1.98	-0.07	5.12	0.04
34	0.00	0.01	0.01	-0.02
35	4.52	0.13	4.33	-0.08
36	2.24	-0.20	26.50	0.36

*Extended Response

Table 20 (cont). Michigan HSPT in Communication Arts: Reading Tryout, Mantel-Haenszel Statistics
Chi-square & Standardized Mean Difference (SMD)
Reference: White or Male/Focus: Black or Female

Form 3

Item	ethnic chi sq.	ethnic SMD	gender chi sq.	gender SMD
1	0.87	-0.07	0.24	-0.02
2	0.02	0.01	0.48	-0.01
3	0.37	-0.05	1.03	-0.05
4	0.69	-0.04	0.70	-0.02
5	9.20	0.17	1.13	0.03
6	0.00	-0.00	0.67	0.04
7	0.06	0.02	6.27	0.01
8	0.27	0.04	0.43	0.02
9	0.35	-0.02	1.13	0.02
10	0.00	-0.01	0.03	-0.00
11	1.75	-0.04	0.00	0.00
12	3.82	-0.12	6.23	-0.08
13	0.00	0.02	0.40	0.01
14	0.16	-0.02	0.01	0.01
15	0.52	-0.04	1.45	-0.04
16	0.46	-0.04	2.38	0.04
17	0.18	0.02	0.13	-0.01
18	2.43	-0.07	1.02	-0.03
19	3.00	-0.09	3.34	-0.06
20	0.02	-0.01	4.16	-0.07
21	0.31	0.05	1.49	-0.04
22	3.03	0.10	1.14	-0.03
23	1.33	0.04	2.71	0.03
24	0.00	0.00	0.94	-0.01
25	1.73	0.05	4.81	0.06
26	0.52	0.04	1.05	0.03
27	0.53	-0.02	4.73	-0.07
28	0.41	-0.02	0.01	-0.00
29	0.01	-0.02	2.20	-0.05
30	3.09	0.09	0.00	0.01
31	0.62	0.05	0.16	0.03
32	0.58	0.06	0.02	0.01
33	1.64	0.05	3.64	0.07
34	3.67	0.10	1.64	0.03
35	0.31	0.01	0.00	0.00
36*	0.44	-0.05	11.56	0.20

Form 4

Item	ethnic chi sq.	ethnic SMD	gender chi sq.	gender SMD
1	7.71	-0.12	1.41	-0.04
2	1.30	0.05	6.04	-0.07
3	0.00	-0.01	0.20	0.01
4	1.70	0.07	19.94	-0.14
5	17.91	0.21	0.62	0.03
6	0.53	-0.03	1.41	0.04
7	0.02	0.01	1.90	0.02
8	2.04	0.07	8.33	-0.08
9	0.50	0.05	5.57	-0.07
10	0.00	-0.01	0.23	-0.01
11	0.87	-0.03	8.84	0.06
12	1.55	-0.06	5.57	0.06
13	0.06	-0.01	3.75	0.03
14	0.22	-0.02	0.73	0.01
15	11.76	-0.14	0.00	0.00
16	0.10	0.02	1.51	0.04
17	0.05	0.01	0.00	-0.00
18	0.40	-0.04	0.06	-0.01
19	0.00	-0.00	3.37	-0.07
20	0.50	0.02	3.19	0.03
21	0.18	0.02	2.44	-0.04
22	1.98	-0.06	1.94	0.03
23	0.00	-0.01	1.40	0.03
24	3.28	0.07	1.66	-0.03
25	0.21	0.02	3.86	0.06
26	0.10	0.00	0.55	0.02
27	0.00	-0.02	8.02	-0.09
28	0.90	-0.04	0.00	0.01
29	0.10	-0.03	2.40	-0.05
30	0.93	-0.04	2.91	-0.04
31	1.49	0.06	0.45	-0.02
32	1.25	0.06	0.01	0.00
33	14.85	0.16	0.34	-0.03
34	0.00	0.00	0.15	-0.01
35	0.67	0.04	0.95	-0.02
36	2.47	-0.12	50.40	0.40

*Extended Response

Table 20 (cont). Michigan HSPT in Communication Arts: Reading Tryout, Mantel-Haenszel Statistics
Chi-square & Standardized Mean Difference (SMD)
Reference: White or Male/Focus: Black or Female

Form 5

Item	ethnic chi sq.	ethnic SMD	gender chi sq.	gender SMD
1	0.09	0.04	4.98	-0.08
2	0.00	-0.01	1.13	0.04
3	0.90	-0.05	1.86	-0.04
4	0.26	-0.03	0.22	0.01
5	0.61	0.04	0.00	-0.00
6	0.00	0.01	0.41	0.01
7	0.91	0.07	1.35	-0.04
8	1.05	-0.05	5.93	0.05
9	0.74	-0.05	0.01	-0.02
10	0.10	0.04	0.12	0.01
11	0.81	0.05	0.36	-0.02
12	3.06	0.12	0.19	0.01
13	1.18	0.07	2.02	-0.04
14	0.32	0.05	1.48	0.04
15	0.93	0.07	1.94	0.05
16	0.05	-0.03	2.97	-0.06
17	0.43	-0.04	0.24	-0.03
18	1.33	-0.09	0.44	-0.02
19	0.20	0.04	2.89	-0.06
20	0.13	-0.02	0.00	-0.01
21	0.10	-0.02	0.21	0.01
22	1.45	0.08	0.93	0.04
23	0.33	0.04	0.22	-0.01
24	0.01	0.02	1.96	-0.05
25	0.26	0.03	1.43	-0.03
26	0.22	0.04	0.90	-0.04
27	0.27	-0.04	2.28	0.04
28	2.21	-0.09	1.32	0.03
29	0.35	0.05	6.81	0.09
30	2.68	0.11	0.00	-0.00
31	1.24	-0.09	2.79	0.07
32	2.82	0.11	0.82	0.04
33	0.09	-0.03	3.77	0.06
34	1.57	0.08	2.65	-0.05
35	0.18	-0.04	0.03	-0.00
36*	2.03	-0.14	5.74	0.17

Form 6

Item	ethnic chi sq.	ethnic SMD	gender chi sq.	gender SMD
1	3.91	-0.11	3.86	-0.07
2	0.93	-0.06	0.03	-0.00
3	2.24	-0.09	1.69	-0.03
4	1.93	0.06	14.87	0.07
5	2.19	-0.07	0.00	-0.01
6	3.25	0.12	3.70	0.07
7	0.10	-0.01	21.71	-0.15
8	0.14	0.02	4.40	0.04
9	0.58	-0.06	0.00	-0.01
10	0.61	0.04	0.91	0.01
11	1.08	-0.05	14.58	0.11
12	1.94	0.09	0.06	0.00
13	0.20	0.04	0.07	0.01
14	0.32	0.04	0.02	-0.00
15	1.39	0.06	0.56	0.01
16	0.06	0.02	0.90	-0.01
17	0.07	-0.03	1.77	0.04
18	5.25	-0.14	0.46	-0.03
19	1.83	0.09	0.01	-0.01
20	1.40	-0.07	2.36	-0.03
21	0.34	-0.04	6.30	0.05
22	5.68	0.15	0.06	0.01
23	1.53	-0.08	2.40	-0.05
24	1.72	-0.08	7.64	-0.08
25	0.15	0.03	6.04	-0.08
26	0.27	-0.03	5.18	-0.08
27	0.77	0.05	0.09	0.02
28	0.01	0.02	7.26	-0.09
29	0.11	0.03	0.41	0.02
30	0.17	0.03	1.95	0.04
31	0.02	-0.00	0.49	0.02
32	0.26	0.04	0.09	-0.01
33	0.06	-0.02	1.18	0.03
34	2.67	-0.10	3.00	-0.05
35	0.72	0.03	0.00	0.00
36	0.78	0.08	20.81	0.31

*Extended Response

Table 20 (cont). Michigan HSPT in Communication Arts: Reading Tryout, Mantel-Haenszel Statistics
Chi-square & Standardized Mean Difference (SMD)
Reference: White or Male/Focus: Black or Female

Form 7

Item	ethnic chi sq.	ethnic SMD	gender chi sq.	gender SMD
1	0.09	0.04	4.98	-0.08
2	0.00	-0.01	1.13	0.04
3	0.90	-0.05	1.86	-0.04
4	0.26	-0.03	0.22	0.01
5	0.61	0.04	0.00	-0.00
6	0.00	0.01	0.41	0.01
7	0.91	0.07	1.35	-0.04
8	1.05	-0.05	5.93	0.05
9	0.74	-0.05	0.01	-0.02
10	0.10	0.04	0.12	0.01
11	0.81	0.05	0.36	-0.02
12	3.06	0.12	0.19	0.01
13	1.18	0.07	2.02	-0.04
14	0.32	0.05	1.48	0.04
15	0.93	0.07	1.94	0.05
16	0.05	-0.03	2.97	-0.06
17	0.43	-0.04	0.24	-0.03
18	1.33	-0.09	0.44	-0.02
19	0.20	0.04	2.89	-0.06
20	0.13	-0.02	0.00	-0.01
21	0.10	-0.02	0.21	0.01
22	1.45	0.08	0.93	0.04
23	0.33	0.04	0.22	-0.01
24	0.01	0.02	1.96	-0.05
25	0.26	0.03	1.43	-0.03
26	0.22	0.04	0.90	-0.04
27	0.27	-0.04	2.28	0.04
28	2.21	-0.09	1.32	0.03
29	0.35	0.05	6.81	0.09
30	2.68	0.11	0.00	-0.00
31	1.24	-0.09	2.79	0.07
32	2.82	0.11	0.82	0.04
33	0.09	-0.03	3.77	0.06
34	1.57	0.08	2.65	-0.05
35	0.18	-0.04	0.03	-0.00
36*	2.03	-0.14	5.74	0.17

Form 8

Item	ethnic chi sq.	ethnic SMD	gender chi sq.	gender SMD
1	3.91	-0.11	3.86	-0.07
2	0.93	-0.06	0.03	-0.00
3	2.24	-0.09	1.69	-0.03
4	1.93	0.06	14.87	0.07
5	2.19	-0.07	0.00	-0.01
6	3.25	0.12	3.70	0.07
7	0.10	-0.01	21.71	-0.15
8	0.14	0.02	4.40	0.04
9	0.58	-0.06	0.00	-0.01
10	0.61	0.04	0.91	0.01
11	1.08	-0.05	14.58	0.11
12	1.94	0.09	0.06	0.00
13	0.20	0.04	0.07	0.01
14	0.32	0.04	0.02	-0.00
15	1.39	0.06	0.56	0.01
16	0.06	0.02	0.90	-0.01
17	0.07	-0.03	1.77	0.04
18	5.25	-0.14	0.46	-0.03
19	1.83	0.09	0.01	-0.01
20	1.40	-0.07	2.36	-0.03
21	0.34	-0.04	6.30	0.05
22	5.68	0.15	0.06	0.01
23	1.53	-0.08	2.40	-0.05
24	1.72	-0.08	7.64	-0.08
25	0.15	0.03	6.04	-0.08
26	0.27	-0.03	5.18	-0.08
27	0.77	0.05	0.09	0.02
28	0.01	0.02	7.26	-0.09
29	0.11	0.03	0.41	0.02
30	0.17	0.03	1.95	0.04
31	0.02	-0.00	0.49	0.02
32	0.26	0.04	0.09	-0.01
33	0.06	-0.02	1.18	0.03
34	2.67	-0.10	3.00	-0.05
35	0.72	0.03	0.00	0.00
36	0.78	0.08	20.81	0.31

*Extended Response

Table 20 (cont). Michigan HSPT in Communication Arts: Reading Tryout, Mantel-Haenszel Statistics
Chi-square & Standardized Mean Difference (SMD)
Reference: White or Male/Focus: Black or Female

Form 9

Item	ethnic chi sq.	ethnic SMD	gender chi sq.	gender SMD
1	13.34	-0.16	4.27	-0.06
2	1.09	0.05	0.54	-0.02
3	1.13	-0.05	0.22	-0.02
4	2.67	0.05	0.93	0.01
5	0.00	-0.00	0.28	0.01
6	0.03	0.01	2.60	0.05
7	0.56	-0.03	1.45	-0.04
8	0.09	-0.02	0.13	-0.02
9	0.11	-0.01	0.77	-0.03
10	0.50	-0.03	0.37	-0.02
11	9.71	0.14	1.27	-0.03
12	0.80	0.04	1.18	0.02
13	1.77	0.06	5.49	0.07
14	1.96	-0.05	8.04	0.05
15	0.49	0.03	11.08	-0.09
16	0.09	-0.02	3.67	-0.04
17	0.12	-0.01	3.96	-0.05
18	0.86	0.04	1.17	0.02
19	0.03	0.01	1.50	-0.03
20	2.05	-0.06	0.20	0.02
21	0.90	0.05	1.10	0.04
22	0.00	-0.00	4.66	0.03
23	1.31	-0.05	1.03	-0.03
24	0.02	-0.01	0.08	-0.01
25	1.62	-0.07	3.38	-0.05
26	5.60	-0.11	0.28	0.02
27	2.98	0.07	3.80	-0.05
28	3.38	0.07	2.22	-0.04
29	0.00	0.01	1.04	0.03
30	4.07	0.07	0.00	0.00
31	0.01	0.01	0.00	-0.00
32	0.03	-0.01	0.35	0.02
33	0.01	-0.00	1.64	0.02
34	1.76	-0.08	0.88	-0.02
35	2.20	0.06	1.83	-0.04
36*	0.96	0.09	27.53	0.33

Form 10

Item	ethnic chi sq.	ethnic SMD	gender chi sq.	gender SMD
1	2.55	-0.10	17.83	-0.12
2	2.45	-0.08	0.00	0.00
3	1.10	-0.09	0.61	-0.03
4	0.01	-0.02	1.40	0.02
5	0.00	-0.02	3.21	0.04
6	0.02	-0.00	1.28	0.04
7	4.53	0.14	0.01	-0.01
8	0.00	-0.02	1.97	-0.06
9	0.03	-0.01	2.88	-0.07
10	0.02	-0.01	3.16	-0.07
11	0.03	0.03	1.25	0.04
12	0.04	-0.00	1.87	0.03
13	0.42	0.04	1.52	-0.04
14	0.60	-0.04	1.41	0.03
15	0.09	0.02	10.12	-0.09
16	0.05	0.03	7.51	-0.08
17	0.50	-0.05	0.01	-0.01
18	2.49	0.09	6.38	-0.09
19	1.98	0.09	1.64	-0.02
20	0.00	0.01	0.01	0.01
21	1.32	0.07	7.95	0.09
22	0.00	0.02	2.98	0.02
23	6.03	-0.16	2.60	0.05
24	0.61	0.04	0.53	0.02
25	0.01	-0.02	0.19	0.02
26	0.27	-0.04	0.22	0.02
27	0.05	0.01	0.01	-0.01
28	0.03	-0.00	4.98	-0.06
29	1.47	-0.08	1.00	0.04
30	0.32	-0.04	0.26	0.02
31	0.24	0.04	0.07	-0.01
32	0.21	-0.02	7.97	0.08
33	0.02	0.03	5.59	0.05
34	0.19	-0.04	2.04	-0.04
35	0.05	0.01	2.85	-0.06
36	2.18	0.14	19.81	0.23

*Extended Response

Michigan High School Proficiency Test
Communication Arts: Reading Tryout
Teacher Comment Sheet

As part of the Michigan HSPT Communication Arts: Reading tryout, the Michigan Department of Education is asking you to complete the following comment sheet.

Directions: Please answer each of the following to the **BEST** of your ability. Each item can be answered by the person administering the HSPT Communication Arts: Reading tryout. None of the items are specific to any particular form. IF YOU NEED MORE SPACE TO RESPOND, PLEASE USE THE BACK OF THESE SHEETS OR ATTACH YOUR OWN.

1. Was the Administration Manual clear, easy to use, and complete? _____Yes _____No
   If "No," what changes would you suggest?

_____

_____

_____

_____

2. Did you have a sufficient number of test materials? _____Yes _____No
   If "No," which ones were insufficient?_____

_____

3. Within the time permitted, approximately what percentage of your students finished:

   Part I_____% Part II_____% Part III_____%

4. Did your students understand the "Directions" for each part of the test? Please estimate the proportion of students who were able to follow the Directions for each part.

   Part I_____% Part II_____% Part III_____%

5. Enter the number of students that read through the reading selections and appeared to be seriously engaged with the test. _____

6. Enter the number of students that appeared NOT to have read the selections completely and were only superficially engaged with the test. _____

   (NOTE: 5 & 6 need not add up to the total number of students tested.)

Michigan High School Proficiency Test
Communication Arts: Reading Tryout
Teacher Comment Sheet (cont'd.)

7. Were the reading selections of appropriate difficulty for grade 11 students? Please comment.

8. Were there particular questions in any part of the test on which a large number of students had difficulty? If so, please indicate either Part I or Part II, and give specifics below.

Part of Test	Item #	Comments

9. What comments, concerns, or issues did students raise about the Part III: Response to the Communication Arts: Reading Selections?

10. Were there other aspects of this test which gave the students or you, as test administrator, difficulty? Please explain.

11. In this section, provide your ideas, critique, etc., on this tryout. Please include student reactions to exercises as well as your overview of the entire test.

THANK YOU FOR YOUR TIME AND EFFORT IN RESPONDING TO THESE QUESTIONS.

Appendix C

Table 22. Michigan HSPT in Communication Arts: Reading Pilot
Descriptive Statistics by Form

Form	Set of Pilot Group	# of Scored Items	# of Points	Mean	s.d.	N	α	P-Value [1] Mean	P-Value [1] s.d.	Item-Test Correlation Mean	Item-Test Correlation s.d.
1	-	36	38	23.60	7.07	1505	.85	.62	.18	.41	.12
	1	-		23.99	6.93	760	-	-	-	-	-
	3	-		23.26	7.03	742	-	-	-	-	-
2	-	36	38	21.96	7.23	1448	.86	.58	.19	.42	.11
	1	-		22.67	7.15	737	-	-	-	-	-
	2	-		21.22	7.24	711	-	-	-	-	-
3	-	36	38	21.22	7.63	1396	.87	.56	.20	.43	.10
	2	-		20.43	7.12	662	-	-	-	-	-
	3	-		21.93	7.47	734	-	-	-	-	-

[1] Includes p-value for extended-response items obtained by dividing the average score by the maximum number of points.

Table 25. HSPT in Communication Arts: Reading Pilot
Frequency of Interrater Agreement for Extended-Response Items by Form

Agreement between first 2 readers: 1 = agree 3 = nonadjacent
2 = adjacent . = student's response invalid

Form 1

Item 36	Frequency	Percent	Cumulative Frequency	Cumulative Percent
missing	201	.	.	.
1	922	68.0	922	68.0
2	341	25.1	1263	93.1
3	93	6.9	1356	100.0

Form 2

Item 36	Frequency	Percent	Cumulative Frequency	Cumulative Percent
missing	244	.	.	.
1	890	67.5	890	67.5
2	345	26.2	1235	93.7
3	83	6.3	1318	100.0

Form 3

Item 36	Frequency	Percent	Cumulative Frequency	Cumulative Percent
missing	260	.	.	.
1	966	77.0	966	77.0
2	232	18.5	1198	95.5
3	56	4.5	1254	100.0

All Forms Together

Item 36	Frequency	Percent	Cumulative Frequency	Cumulative Percent
missing	705	.	.	.
1	2778	70.7	2778	70.7
2	918	23.4	3696	94.1
3	232	5.9	3928	100.0

Table 27. Michigan HSPT in Communication Arts: Reading Pilot
DIF Statistics (Standardized Mean Differences: SMDs) for Gender and Ethnic Groups

Gender

Form	# of Items	# of Males	# of Females	DIF Against Males SMD≥.20	DIF Against Males .19≥SMD≥.10		DIF Against Females SMD≤-.20	DIF Against Females -.19≤SMD≤-.10
1	36	724	769	0	1	(0)*	0	1
2	36	709	726	0	1	(1)	0	0
3	36	686	698	0	3	(0)	0	3

Ethnicity

Form	# of Items	# of Whites	# of African-Americans	DIF Against Whites SMD≥.20	DIF Against Whites .19≥SMD≥.10		DIF Against African-Americans SMD≤-.20	DIF Against African-Americans -.19≤SMD≤-.10
1	36	1151	173	0	1	(2)*	0	3
2	36	1127	164	0	2	(2)	1	2
3	36	1048	190	0	2	(2)	0	0

* Absolute value of the difference in total "practically significant" DIF across the two groups of a comparison. Total DIF for each group is twice the number of items with |SMD|≥.20 plus the number of items with .10≤|SMD|≤.19.

Appendix D

Table 30. Student Survey in Communication Arts: Reading

Directions: *Listed below are statements about activities that often take place as a part of school experiences. The Michigan Department of Education is interested in finding out how often these activities have been a part of <u>your</u> school experience by the end of tenth grade.*

Please read each question carefully and answer it the BEST that you can. For each question, darken one circle on your answer sheet labeled Session 1 to indicate your response using the scale below.

Scale	A	B	C	D
	Never	Very Little	Some	A Lot

Sample Item:

By the end of tenth grade, how often did your school experience include:

A. reading comic books?

By the end of tenth grade, how often did your school experience include:

1. reading short stories?
2. reading essays?
3. reading plays?
4. reading speeches?
5. reading poems?
6. reading magazine articles?
7. reading newspaper articles?
8. reading documents (historical and legal)?
9. reading editorials?
10. reading editorial cartoons, political cartoons, graphs, charts, etc.?
11. reading technical manuals (computer manuals, car repair manuals, etc.)?
12. reading a variety of reading material based on the same theme or main idea?
13. applying ideas from a variety of reading materials that relate to the same current issue?
14. responding to reading selections by writing a one- to two-page essay taking a position on an issue and supporting it from what you have read?
15. reading silently for an extended period of time (30-40 minutes) in class?

By the end of tenth grade, how often did your school experience include:

16. answering questions that relate to one reading selection?
17. answering questions that relate to two or more reading selections?
18. connecting information and ideas from more than one reading selection?
19. evaluating and reacting critically to reading materials that you have read?
20. identifying the central purpose and/or the theme from a variety of types of reading material?
21. demonstrating your knowledge of different text features (e.g., graphs, marginal notations, headings, subtitles, etc.)?
22. demonstrating your knowledge of use of different literary devices (e.g., foreshadowing, flashback, etc.)?
23. the independent use of a variety of reading strategies (e.g., mapping, summarizing, note taking, Directed Reading and Thinking Activities (DRTA), Survey-Question-Read-Recite-Review (SQ3R), etc.)?

Thank you very much!

Table 31. Student Survey Response Means in Reading
("*" Means greater than 10% of students responded "never".)
(sorted by mean value)

By the end of tenth grade, how often did your school experience include:

Statement #	Statement	Mean
11*	reading technical manuals (computer manuals, car repair manuals, etc.)?	.79
9*	reading editorials?	.96
8*	reading documents (historical and legal)?	1.16
4*	reading speeches?	1.22
10*	reading editorial cartoons, political cartoons, graphs, charts, etc.?	1.48
7*	reading newspaper articles?	1.54
6*	reading magazine articles?	1.61
3	reading plays?	1.61
2	reading essays?	1.62
21*	demonstrating your knowledge of different text features (e.g., graphs, marginal notations, headings, subtitles, etc.)?	1.63
13	applying ideas from a variety of reading materials that relate to the same current issue?	1.73
14*	responding to reading selections by writing a one-to two-page essay taking a position on an issue and supporting it from what you have read?	1.78
19	evaluating and reacting critically to reading materials that you have read?	1.82
23*	the independent use of a variety of reading strategies (e.g., mapping, summarizing, note taking, Directed Reading and Thinking Activities (DRTA),Survey-Question-Read-Recite-Review (SQ3R), etc.)?	1.82
22	demonstrating your knowledge of use of different literary devices (e.g., foreshadowing, flashback, etc.)?	1.86
18	connecting information and ideas from more than one reading selection?	1.88

12	reading a variety of reading material based on the same theme or main idea?	1.94
5	reading poems?	1.97
20	identifying the central purpose and/or the theme from a variety of types of reading material?	2.05
17	answering questions that relate to two or more reading selections?	2.05
15	reading silently for an extended period of time (30-40 minutes) in class?	2.06
1	reading short stories?	2.34
16	answering questions that relate to one reading selection?	2.51

Table 32. Student Survey Mean Scores by Part
(0 = "never" to 3 = "a lot")

Part	Mean
Construct Meaning	2.01
Knowledge about Reading	1.77
Cross-Text	1.91
Response to the Reading Selection	1.87
General (Questions applying to all parts)	1.53

Table 33. Teacher Survey - Communication Arts:
Reading Statements with ≥20% Schools Responding NT

(N=249)

Statement	% of Schools Responding NT
35	51%
32	30%
33	21%
28	20%

Types of Reading (Genre):

35. Reading technical manuals (computer manuals, car repair manuals, etc.)

32. Reading documents (historical and legal)

33. Reading editorials

28. Reading speeches

Statements with ≥50% Schools Responding NSI

NA. The highest percentage is 27% (N=67) for Statement 37.

37. (Part G. Objectives) Integrating and applying ideas from a variety of reading materials that relate to the same current issue

MICHIGAN HIGH SCHOOL PROFICIENCY TEST IN
COMMUNICATION ARTS: READING

Tryout and Pilot Technical Report Development Team

(alphabetically)

Jane K. Faulds

Catherine B. Smith

Jean W. Yan

Correspondence concerning this report should be addressed to:

Jean Yan or Catherine Smith
MEAP Office
Michigan Department of Education
P.O. Box 30008
Lansing, MI 48909
(517) 373-8393 (o)
(517) 335-1186 (fax)
yanj@state.mi.us
smithcb@state.mi.us

This development team wishes to thank the following people for their time and expertise in reviewing this document and providing suggestions and comments: Mr. Vern Yetman and Ms. Pamela Crosby. CTB/McGraw-Hill, as contractor for the development phase of the HSPT in Mathematics, Reading and Science, provided data and most statistical analyses used in this report.

Carolyn Male and Yolanda Y. Stephens
provided excellent support services for this project.